Theor Appl Genet (2009) 119:417–427
DOI 10.1007/s00122-009-1049-7

ORIGINAL PAPER

QTL in mega-environments: I. Universal and specific seed yield QTL detected in a population derived from a cross of high-yielding adapted × high-yielding exotic soybean lines

Laura Palomeque · Liu Li-Jun · Wenbin Li · Bradley Hedges · Elroy R. Cober · Istvan Rajcan

Received: 10 November 2008 / Accepted: 21 April 2009 / Published online: 22 May 2009

Abstract Modern soybean [(*Glycine max* (L.) Merrill] breeding programs rely primarily on the use of elite × elite line crosses to develop high-yielding cultivars. Favorable alleles for traits of interest have been found in exotic germplasm but the successful introduction of such alleles has been hampered by the lack of adaptation of the exotic parent to local mega-environment and difficulties in identifying superior progeny from elite × exotic crosses. The objective of this study was to use a population derived from a cross between an adapted and an exotic elite line to understand the genetic causes underlying adaptation to two mega-environments (China and Canada). A cross between a high-yielding Canadian cultivar 'OAC Millennium' and an elite Chinese cultivar 'Heinong 38' was performed to develop a recombinant inbred line (RIL) population. The RIL population was evaluated in China and Canada in multiple environments from 2004 to 2006. Significant variation for seed yield was observed among the RILs in both the Chinese and Canadian environment. Individual RILs performed differently between the Chinese and Canadian environments suggesting differential adaptation to intercontinental mega-environments. Seven seed yield quantitative trait loci (QTL) were identified of which five were mega-environment universal QTL (linked to markers Satt100, Satt162, Satt277, Sat_126, and the interval of Satt139-Sat_042) and two were mega-environment-specific QTL (at marker intervals, Satt194-SOYGPA and Satt259-Satt576). Seed yield QTL located near Satt277 has been confirmed and new QTL have been identified explaining between 9 and 37% of the phenotypic variation in seed yield. The QTL located near Satt100 explained the greatest amount of variation ranging from 18 to 37% per environment. Broad sense heritability ranged from 89 to 64% among environments. Epistatic effects have been identified in both mega-environments with pairs of markers explaining between 9 and 14% of the phenotypic variation in seed yield. An improved understanding of the type of QTL action as either universal or mega-environment-specific QTL as well as their interaction may facilitate the development of strategies to introgress specific high-yielding alleles from Chinese to North American germplasm and vice versa to sustain efforts in breeding of high-yielding soybean cultivars.

Communicated by F. Muehlbauer.

L. Palomeque · I. Rajcan (✉)
Department of Plant Agriculture, Crop Science Building, University of Guelph, 50 Stone Road East, Guelph, ON N1G 2W1, Canada
e-mail: irajcan@uoguelph.ca

L. Li-Jun
Soybean Research Institute of Heilongjiang, Harbin, Heilongjiang, People's Republic of China

W. Li
Soybean Research Institute of the Northeast Agricultural University, Harbin, Heilongiang, People's Republic of China

B. Hedges
Pioneer Hi-bred Ltd, A DuPont Company, Chatham, ON, Canada

E. R. Cober
Agriculture and Agrifood Canada, Eastern Cereal and Oilseed Crop Research Centre, Ottawa, ON, Canada

Introduction

Most modern soybean breeding programs rely almost exclusively on the use of elite lines in crosses to develop

high-yielding cultivars. However, pedigree analysis showed that 80% of the North American germplasm ancestry can be traced back to only 13 ancestral cultivars (Gizlice et al. 1994). Exotic germplasm within the soybean [(*Glycine max* (L.) Merrill] gene pool is a rich source of many potentially favorable alleles that could improve seed yield. Most of the yield improvement in soybean has been made by using crosses among elite germplasm rather than crosses with non-adapted cultivars or wild relatives (Carter et al. 2004). The use of plant introductions (PI) has been hampered by the difficulty in obtaining similar or higher yield in the selected progeny when compared with the progeny obtained using only adapted parents (Smalley et al. 2004). It has been argued that the lack of success in using PIs as parents is related to yield being a complex quantitative trait and the linkage of favorable yield alleles to unfavorable ones and the presence of deleterious alleles in much of the progeny (Concibido et al. 2003). Another reason could be that the elite lines may have a greater number of favorable alleles fixed compared to the PI (Bernardo 2002).

Mega-environments have been defined by CIMMYT as "a broad, not necessarily contiguous area, occurring in more than one country and frequently transcontinental, defined by similar biotic and abiotic stresses, cropping system requirements, consumer preferences, and, for convenience, by a volume of production" (Braun et al. 1996). As a quantitative trait, seed yield is affected by the environment and needs to be measured in several environments to determine its genetic base (Orf et al. 1999a). Most often seed yield quantitative trait loci (QTL) are identified using the average seed yield measured across environments, even though individual environmental data are analyzed from such diverse places as North and South American mega-environments (Orf et al. 1999a). Using two populations of recombinant inbred lines (RILs) derived from the crosses 'Minsoy' × 'Archer', and "Noir 1' × 'Archer', Orf et al. (1999a) identified QTL for plant height, lodging, days to flowering, maturity, reproductive period, seed yield, seed weight, and seed oil and protein concentration. In many cases, the beneficial alleles came from 'Minsoy' and 'Noir 1', which were the exotic parents in crosses with the North American-adapted cultivar 'Archer' (Mansur et al. 1996). Kabelka et al. (2004) reported 15 QTL for seed yield using a population derived from a cross between 'BSR 101' (adapted) and LG82-8379 (PI-derived breeding line), for which nine of the QTL carried a high-yielding allele from the PI-derived parent. Using three backcross-derived populations of soybean that involved PIs as donor parents, Guzman et al. (2007) confirmed 13 seed yield QTL for which in 9 QTL the beneficial allele was contributed by the PI parent. Seed yield QTL were identified in a soybean population developed from a cross between a PI and an elite line where both parental lines were selected for high yield in the same environment (Smalley et al. 2004). Zhao et al. (2005) evaluated a population obtained from a cross between a European and a Chinese rapeseed (*Brassica napus*) in two different countries and found 18 QTL associated with high oil content. The authors argued that the reason for obtaining more than twice the expected number of QTL associated with the trait of interest was attributed to the use of two parents coming from two different breeding programs, one from China and the other from Germany (Zhao et al. 2005). None of the above studies were conducted in the environment where the PI or exotic parent came from.

The relatively low success rate of obtaining PI-derived high-yielding lines in soybean could be due to a limited understanding of how to properly evaluate the results obtained from progenies that are the product of crossing a PI with an elite parent. Should the product of a cross between an elite soybean line and a PI be only evaluated where the elite line is adapted? Or, should the progeny be evaluated in the mega-environments of adaptation of both parental lines? If the progeny is evaluated in both mega-environments, would QTL associated with the trait of interest behave as mega-environment universal QTL (QTL_U) or as mega-environment-specific QTL (QTL_{SP})? If these QTL_U and QTL_{SP} exist and breeders want to include exotic germplasm in their breeding germplasm pool, they would have to search for PIs that have the favorable alleles at the QTL_U and/or QTL_{SP} for the trait of interest. A better understanding of the relative contribution of mega-environment universal and specific QTL will help breeders in the evaluation and identification of the suitable PI as a source of parents for generating superior breeding populations maximizing the use of exotic alleles. To understand the relationship between mega-environment universal and specific QTL, it is important to dissect and analyze the environmental effects on individual and interacting QTL, which could not be done by using the mean environment.

High yield can result from QTL that interact with other QTL (Lark et al. 1995). It is, therefore, important to evaluate the relationship between seed yield QTL and QTL associated with other traits of interest. If the epistatic QTL are not "transferred" together with the yield QTL to the cultivar that is being developed, yield improvement will be unsuccessful because high yield is conditional on the presence of epistatic effects (Lark et al. 1995). Therefore, the objectives of this study were: (1) to determine whether or not evaluation of a mapping population in different mega-environments results only in mega-environments-specific, only universal QTL, or both; (2) to estimate the impact of epistasis on the potential of developing high-yielding progeny from crosses involving PIs when evaluated in reciprocal mega-environments.

Materials and methods

Experimental design

A population consisting of 98 $F_{4:7}$ RIL developed from the cross OAC Millennium × Heinong #38 was used in this study. Five lines that were contaminated were not included in the analysis, therefore only 93 RIL were analyzed. F_1's were produced by crossing OAC Millennium with Heinong #38. Both parents had indeterminate growth habit. OAC Millennium [2,650 crop heat units—CHU (OMAFRA 2002), maturity group 0] is a high yielding, stable cultivar, developed at the University of Guelph and adapted to the Canadian mega-environment. Heinong #38 was developed in the Heilongjiang province of the People's Republic of China and is a well-established cultivar widely grown there. F_2 seeds were advanced by single seed decent for one generation in the greenhouse and the other in Los Andes, Chile to produce F_4 seeds. The F_4 seeds were planted in the field at the Woodstock Research Station in Woodstock, Ontario, where F_4 single plants were harvested individually and threshed separately. No selection was practiced at any generation. The $F_{4:5}$ lines were grown in Woodstock, Ontario and seed was advanced and increased over the next two generations to form $F_{4:7}$ lines, which were grown in yield trials in Canada and China as described below. Parents of the cross and 10 fillers (commercial cultivars from either country) were included in the field trials (Fig. 1).

Field trials were planted at six different locations in two mega-environments (Canada and China) as follows. In 2004, three trials were grown, one at Tavistock (43°18′N 80°49′W) and one at Woodstock (43°7′N 80°45′W) both in Ontario, Canada and a third one at Harbin (45°45′N 126°36′E) in Heilongjiang province in China. The parental line Heinong #38 was not evaluated during this year. In 2005, three trials were planted in Canada, one at Ottawa (45°24′N 75°42′W), one at Tavistock (43°18′N 80°49′W) and one at Woodstock (43°7′N 80°45′W) in Ontario. In 2006, two trials were planted at Harbin (45°45′N 126°36′E) in Heilongjiang, China at two different sites. In each trial, the experiment was arranged as a rectangular lattice design with two replications. The plot size was 1.5 m × 5.5 m with four rows planted at 35 cm between row spacing in the Canadian mega-environment (except in Tavistock where two row plots were planted). In China, the plot sizes were of 2.8 m × 3 m in one site and 2.8 m × 5 m in the other with four elevated ridges per plot, planted at 70 cm between ridges. Seed density was the same in all locations within each mega-environment being at 18 seeds per meter of length in Canada and 28 seeds per meter in China.

Phenotypic scoring, DNA extraction and SSR markers

Seed yield (kg/ha) was measured and adjusted at 13% moisture at each location and year. In 2004, no moisture data was recorded at Harbin. Emergence was recorded in each plot to determine if there was a correlation between emergence and seed yield. No emergence data was recorded in 2006 in either environment in China. Plants were scored on a scale from 0 to 10 in each plot, where 0 corresponded to no emergence and 10–100% emergence.

Tissue samples from parents and RILs were collected in 2004 in Woodstock, Canada, with the aid of a leaf puncher. Leaf disks were collected in pre-labeled 1.5 ml screw-cap tubes. Tubes were transported on ice and then stored at −80°C. The collected tissue was freeze dried with a Labonco FreeZone® freeze dry system (Savant Moduly 0, Kansas City, MO, USA) to remove moisture and inactivate endogenous nucleases and then stored at −80°C. Genomic DNA was obtained from 10 to 15 stored leaf disks per line and extracted with a GeneElute™ Plant Genomic DNA Miniprep Kit (SIGMA®, Saint Louis, MO, USA). A DU-64 Spectrophotometer (Beckman Coulter, Fullerton, CA, USA) was used to quantify the amount of DNA extracted per line. For PCR reactions, the extracted DNA that was collected was diluted in a 1/100 proportion (1 µl DNA in 99 µl deionised and distilled water) and stored as template DNA in a 96-well polypropylene plate (2 ml capacity/well) and kept at 4°C (Primomo et al. 2005). Reagents were combined in the following amounts in a 1.5 ml centrifuge tube or a 15 ml centrifuge tube depending on the number of PCR reactions done each day: 10× PCR buffer, minus Mg 1.5 µl; 50 mM $MgCl_2$; 0.75 µl, Platinum® Taq DNA polymerase, 0.2 µl (all three reagents from Invitrogen™, Carlsbad, CA, USA); dNTP (3 mm) 1 µl, primer forward and reverse 2 µl, each deionised and distilled water 4.55 µl. In a 96-well PCR plate, 12 µl of master mix (combination of reagents previously described) and 3 µl of template DNA were placed per well and spun in a centrifuge (Hermle Z180M, Labnet, Edison, NJ, USA). Finally, 12 µl of mineral oil were added to each well to prevent evaporation. A sealing film was placed over each PCR plate. PCR reactions were performed in the 96-well RoboCycler® (Stratagene, La Jolla, CA, USA) with an amplification program consisting of 2 min at 95°C, followed by 35 cycles of denaturation at 92°C for 1 min, annealing at 47°C for 1.5 min, and extension 68°C during 1.5 min. A final extension step followed at 72°C for 5 min and the completed reaction mixture was held at 4°C. Amplified products were separated by electrophoresis on 5% MetaPhor® agarose (Bio Whittaker Molecular Applications, Rockland, ME, USA) gels using a Sunrise™ 96 Horizontal Electrophoresis Apparatus (Gibco BRL, Life Technologies, Carlsbad, CA, USA) with 115–130 mAmps/volts of current

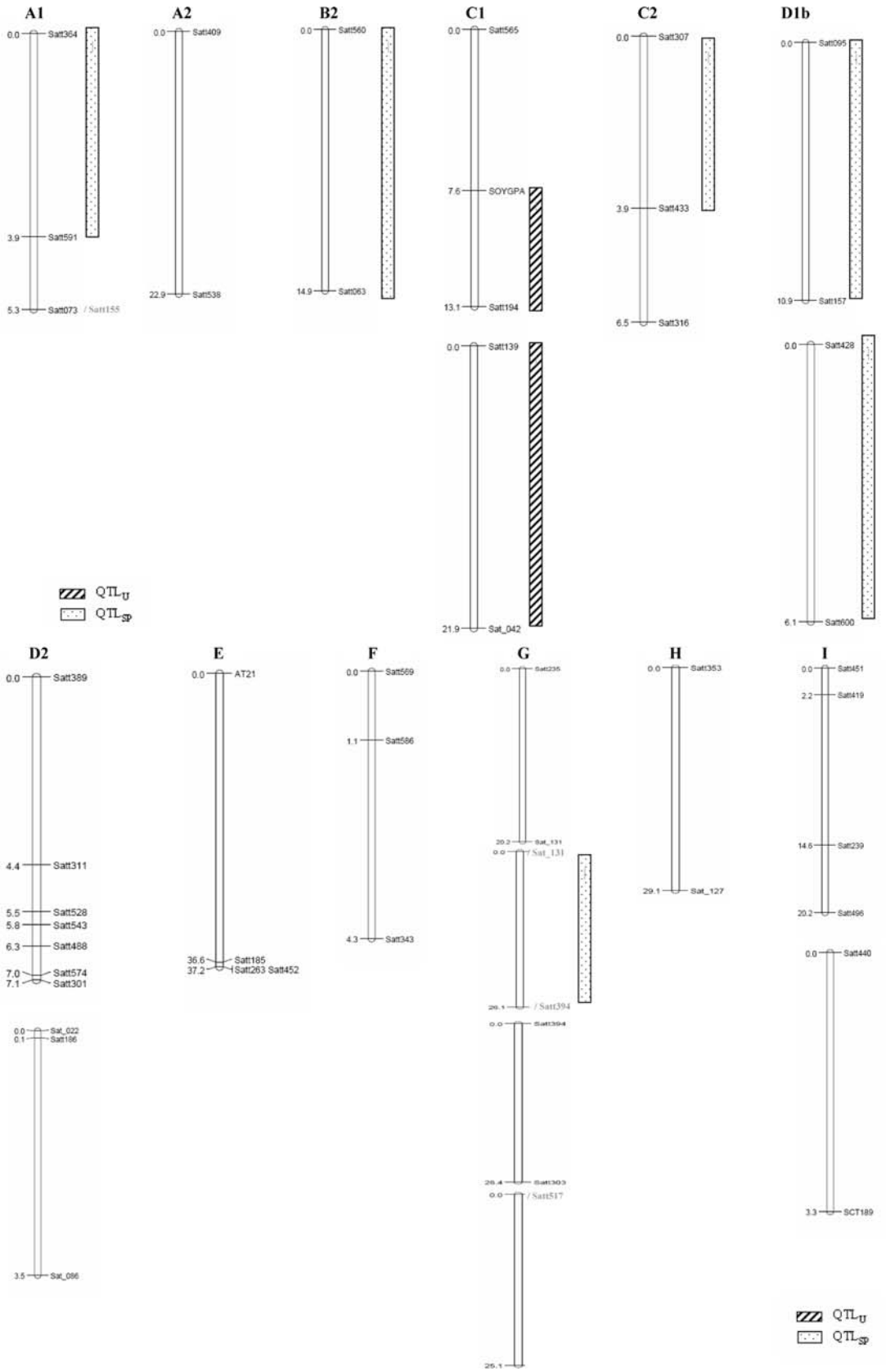

Fig. 1 Partial linkage map developed using the cross 'OAC Millennium × Heinong 38' showing the position of significant mega-environment universal and specific QTL as determined by interval mapping

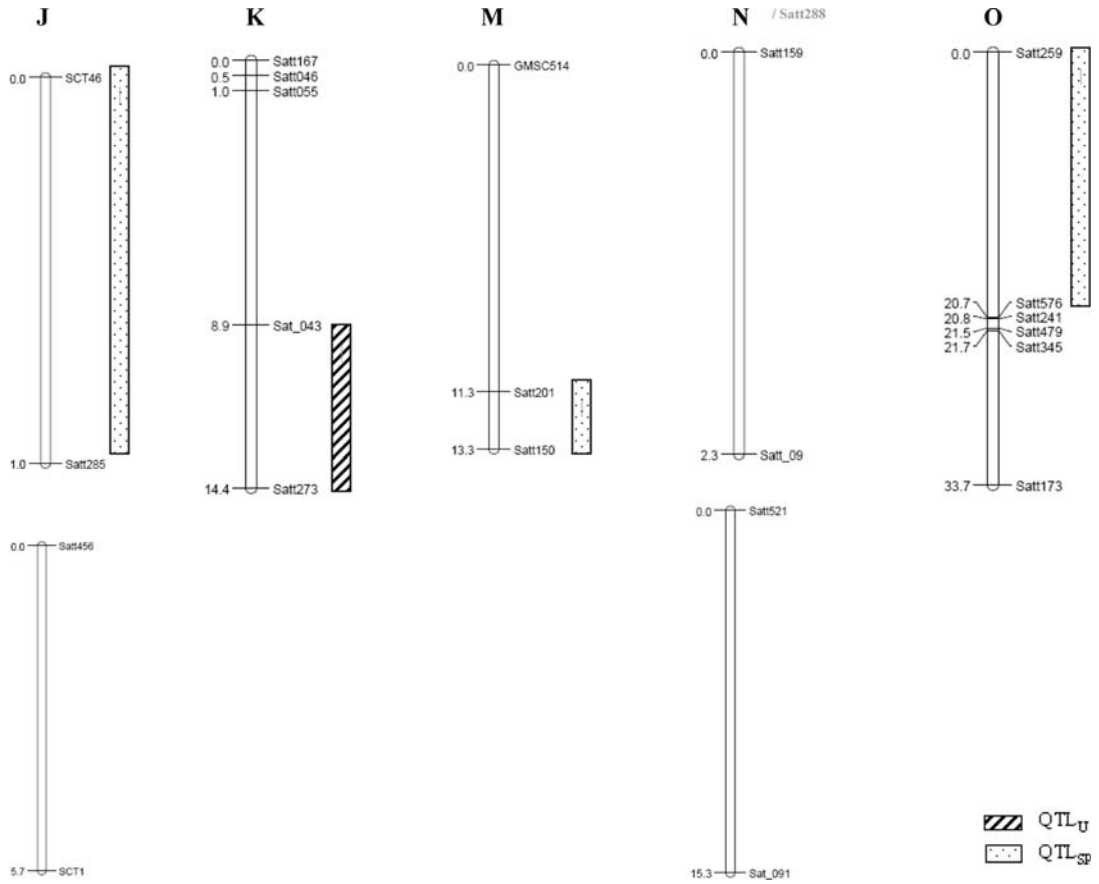

Fig. 1 continued

supplied by an EC105 Electrophoresis Power Supply (ThermoEC, Holbrook, NY, USA). DNA bands were visualized under UV light after staining with ethidium bromide.

The SSR library available at the University of Guelph, ON, Canada, consisted in 450 SSR primer pairs selected from the integrated soybean genetic map (Cregan et al. 1999), which covered approximately equidistantly the 20 soybean genetic linkage groups. Both parental lines were screened with the available primer pairs. One hundred and five of the available 450 SSR were found to be polymorphic between the parents and were used to genotype the entire population.

Linkage mapping and QTL analysis

A linkage map was obtained with JoinMap 4® program (van Oijen 2004). Segregation distortion was calculated with the same program to determine the departure from the expected 1:1 ratio. Eighteen of the polymorphic markers were significantly distorted from the expected ratio, which is in accordance with previous studies (Reinprecht et al. 2006). Markers that showed extreme segregation distortion toward one of the parents and markers that mapped to exactly the same genetic position were excluded from the calculations of the linkage map (van Oijen 2004). The parameter that was used to group markers into LGs was recombinant frequency. The mapping algorithm used was regression and the mapping function selected was Kosambi's. Twenty-six linkage groups were obtained. Marker distance ranged between 0.4 (LG F, K, and O) and 36.6 cM (LG E). The mapping order and distances between markers was found to be similar to the integrated public soybean genetic map (Song et al. 2004). A difference in marker distances was observed in LG G.

Statistical analysis

The variation in seed yield within a test location was partitioned into the effects of genotype, replications and incomplete blocks within replications using the PROC MIXED procedure from SAS ver. 9.1.3 (SAS Institute Inc. 2003) for a rectangular lattice design (Bowley 1999). Blocks within replications were considered to be random variables. Tests of residuals were evaluated using the PROC UNIVARIATE and PROC PLOT procedure SAS ver. 9.1.3 (SAS

Table 1 Least square means and standard error for seed yield of OAC Millennium (Canadian parent) and Heinong #38 (Chinese parent), the recombinant inbreed lines (RIL) derived from them and the 10 lowest- and highest-yielding RIL

Macro-environment		Seed yield means ± SE (kg ha^{-1})				
Location	Year	Parent		RIL population	Mean of 10 highest-yielding RILs	Mean of 10 lowest-yielding RILs
		Canadian	Chinese			
Canada						
Woodstock	2004	2,266 ± 194.9	–	2,066 ± 196.6	2,859 ± 196.2	1,268 ± 198.4
Tavistock	2004	1,934 ± 142.1	–	1,558 ± 137.5	2,213 ± 142.6	952 ± 137.6
Woodstock	2005	4,376 ± 267.7	1,874 ± 267.9	3,701 ± 267.4	4,548 ± 266.1	2,538 ± 266.5
Tavistock	2005	2,966 ± 174.9	1,535 ± 175.2	2,703 ± 176	3,343 ± 176	2,001 ± 176
Ottawa	2005	2,762 ± 124.9	2,094 ± 126	2,628 ± 126.4	3,001 ± 126.1	2,295 ± 125.8
China						
Harbin 1	2004	5,277 ± 500.1	–	4,067 ± 529.9	5,213 ± 539	2,768 ± 541.2
Harbin 1	2006	2,871 ± 322.9	3,434 ± 456.9	2,366 ± 327.5	3,303 ± 323.2	1,298 ± 323
Harbin 2	2006	5,006 ± 960.8	5,697 ± 960.8	4,083 ± 913.3	6,786 ± 913.3	1,990 ± 913.4

Institute Inc. 2003). Seed yield means were not averaged over years and locations since significant interactions were observed between genotypes and years and between genotypes and locations (Reinprecht et al. 2006). Variance components attributable to variation among lines (VG) and residual variation (VE) were derived and used to estimate broad sense heritability for seed yield (VG/(VG + VE) (Huynh et al. 2008). PROC CORR and PROC PLOT procedures were performed to determine the linear correlation between emergence and seed yield. Emergence was included as a covariate in the seed yield model when the correlation was found to be significant. The type I error rate (α) was set at 0.05.

QTL were detected by performing a single-factor ANOVA and multiple QTL mapping (MQM) (Neto et al. 2007). Interval mapping (IM) was performed with MapQTL5® (van Ooijen 2004) with a LOD score of 2.6. The LOD score was calculated by performing a permutation test with MapQTL5® (van Ooijen 2004) with a set of 10,000 iterations. Markers linked to QTL in the IM were used as cofactors in the MQM analysis. Single-factor ANOVA was performed to evaluate unlinked SSR markers and QTL tagged with only one marker (Reinprecht et al. 2006). The association between individual SSR markers and seed yield was determined by using a macro program from SAS ver 9.1.3 (SAS Institute Inc. 2003) obtained from Dr. Elizabeth Lee at the University of Guelph (Guelph, ON, Canada). Epistatic effects and the amount of variation explained by the interaction (R^2) were calculated by EPISTACY 2.0 macro (Holland 1998). The type I error rate (α) for both macros as well as for the segregation distortion analysis was set at 0.01.

Results

Phenotypic evaluation

In the Canadian mega-environment, the mean of the 10 highest-yielding lines had an increase in seed yield when compared with the adapted Canadian parent, OAC Millennium, except at one location (Table 1). At Woodstock in 2005, no differences were found between the adapted parent and the 10 highest-yielding lines. On average, the highest-yielding RILs had an increase in seed yield of 11% compared to the adapted parent. RIL 3 appeared to be the highest-yielding line across macro-environments in Canada, having a seed yield increase of 12% greater than the adapted parent (data not shown). The mean of the 10 lowest-yielding lines yielded less than the adapted parent but on average 24% more than the non-adapted Chinese parent, Heinong #38 in the same macro-environments where the Chinese parent was evaluated (Table 1). The population mean was observed to be inferior to the adapted parent, OAC Millennium. However, on average, it yielded 64% more than the non-adapted parent (Heinong #38) in the same macro-environments in which the Chinese parent was evaluated.

In the Chinese mega-environment, the 10 highest-yielding RILs exceeded the mean seed yield of the adapted parent (Heinong #38, the Chinese elite line) in only one macro-environment. The mean of the 10 highest-yielding RIL exceeded the non-adapted Canadian parent, OAC Millennium (Table 1). RIL 25 was the highest-yielding line in China, with an increase in seed yield of 15% compared to the adapted parent (data not shown). Contrary to what

was found in the Canadian mega-environment, instead of an increase in seed yield, it was observed that the 10 lowest-yielding lines averaged 54% less seed yield than the non-adapted parent (OAC Millennium). The lower seed yield of the 10 lowest-yielding RILs in China could be the result of the fact that the population was developed in the Canadian mega-environment rather than in China.

Broad sense heritability for seed yield was calculated for each environment as follows: for Woodstock 2004 84%, Tavistock 2004 89%, Harbin 2004 64%, Tavistock 2005 85%, Ottawa 2005 79%, Woodstock 2005 81%, Harbin 1 in 2006 78%, and Harbin 2 in 2006 73%.

Mega-environment universal QTL associated with seed yield

Mega-environment universal QTL (QTL_U) was defined as a QTL that was associated with the trait in at least one macro-environment in each of the two mega-environments. A macro-environment corresponded to the combination of year and location (Nyquist 1991). A seed yield QTL_U was detected using the MQM method on LG C1 in the interval tagged by Satt139 and Sat_042 (Table 2). This seed yield QTL_U explained 19% of the variation observed in seed yield.

When single-factor ANOVA was performed using unlinked markers four QTL tagged by four different markers were significantly associated with seed yield in both the Canadian and the Chinese mega-environment (Table 3). Two of the QTL linked each to Satt100 and Satt277, were located on the same LG (C2). The distance between Satt277 and Satt100 is estimated to be about 10 cM in the published consensus map (Song et al. 2004). Each of the QTL_U explained a substantial amount of the phenotypic variance for seed yield, with R^2 ranging between 9 (linked to Satt162) and 37% (linked to Satt100). The QTL_U identified by Satt100 was consistent across environments and accounted for most of the variation in seed yield.

Only two of the five QTL_U have been previously reported (Table 4). Satt277 was associated with lodging, seed weight, plant height, yield per plant height (Orf et al. 1999a), seed yield (Orf et al. 1999b; Orf et al. 1999a; Reinprecht et al. 2006; Smalley et al. 2004; Specht et al. 2001), oil and protein concentration (Reinprecht et al. 2006). Satt162 was reported to be associated with oil and protein concentration (Reinprecht et al. 2006).

Both parents contributed favorable alleles. The Canadian parent contributed the high-seed yield allele in QTL_U linked to Satt162 and Sat_126. The high-seed yield allele came from the Chinese parent in QTL_U detected by Satt100, Satt139 and Satt277 (Table 4).

Mega-environment-specific QTL associated with seed yield

A mega-environment-specific QTL (QTL_{SP}) was defined as the one that was associated with the trait in only one

Table 2 Mega-environment universal and specific seed yield QTL identified by SSR markers found with multiple QTL mapping (MQM) across different environments in Canada and China, from 2004 to 2006

Markers linked to QTL	Linkage group	Woodstock 2004		Tavistock 2004		Tavistock 2005		Harbin (1) 2006	
		R^{2a}	LOD[b]	R^2	LOD	R^2	LOD	R^2	LOD
Satt139-Sat_042	C1	16	2.7	–	–	20	3.63	13	2.77
Satt194-SOYGPA	C1			–	–	–	–	14	3.07
Satt259-Satt576	O	–	–	14	2.9	–	–	–	–

[a] Amount of phenotypic variation in seed yield explained by the QTL

[b] Value on the QTL peak (tagged by the first marker)

Table 3 Mega-environment universal seed yield QTL identified by single-factor ANOVA across different environments in Canada and China, from 2004 to 2006

Marker	Linkage group	Woodstock 2004		Tavistock 2004		Woodstock 2005		Tavistock 2005		Ottawa 2005		Harbin (1) 2006		Harbin (2) 2006	
		R^{2a}	P^b	R^2	P	R^2	P	R^2	P	R^2	P	R^2	P	R^2	P
Satt100	C2	35	0.0001	28	0.0001	–	–	18	0.0002	13	0.0017	19	0.0001	37	0.0001
Satt162	I	–	–	–	–	9	0.0035	12	0.0006	–	–	–	–	12	0.001
Satt277	C2	18	0.0001	18	0.0001	–	–	15	0.0003	–	–	17	0.0001	29	0.0001
Sat_126	K	–	–	–	–	–	–	17	0.0003	–	–	11	0.0062	15	0.001

[a] Amount of phenotypic variation in seed yield explained by the QTL

[b] P value

Table 4 Allelic means and parent contributing favorable allele for the markers linked to mega-environment universal and specific seed yield QTL tested across different environments in Canada and China, from 2004 to 2006

Mega-environment	Location	Year	SSR marker	Allelic means (kg ha^{-1})				Parent allele	Trait associated and previously reported
				A		B			
				Mean	SE	Mean	SE		
Canada	Woodstock	2004	Satt100	1897	47.3	2514	77.9	Chinese	–
			Satt139	1919	59	2281	72.2	Chinese	–
			Satt277	1918	54.2	2349	84	Chinese	Lodging, seed weight, seed yield, plant height, yield/plant height, oil and protein content
	Tavistock	2004	Satt100	1437	38.2	1881	64.7	Chinese	–
			Satt259	1442	47.1	1712	54.4	Chinese	–
			Satt277	1458	42.7	1806	68.1	Chinese	Lodging, seed weight, seed yield, plant height, yield/plant height
	Woodstock	2005	Satt162	3894	87.4	3537	80.9	Canadian	Seed oil and protein concentration
	Tavistock	2005	Satt100	2589	46.9	2972	77.7	Chinese	–
			Satt139	2549	50.5	2921	62.4	Chinese	–
			Satt162	2860	60.1	2566	56.3	Canadian	–
			Satt277	2588	49.9	2945	78	Chinese	Lodging, seed weight, seed yield, plant height, yield/plant height
			Sat_126	2826	52.1	2463	69.4	Canadian	–
	Ottawa	2005	Satt100	2588	24.3	2711	40.2	Chinese	–
China	Harbin 1	2006	Satt100	2204	66.3	2774	109.9	Chinese	–
			Satt139	2191	74.9	2624	92.6	Chinese	–
			Satt194	2191	79.2	2610	87.6	Chinese	–
			Satt277	2200	69.5	2752	108.6	Chinese	Lodge, seed weight, seed yield, plant height, yield/plant height
			Sat_126	2498	76	2096	101.2	Canadian	–
	Harbin 2	2006	Satt100	3523	148.3	5539	256.9	Chinese	–
			Satt162	4627	219.7	3588	209.7	Canadian	–
			Satt277	3532	167.6	5239	272.1	Chinese	Lodging, seed weight, seed yield, plant height, yield/plant height
			Sat_126	4577	198.6	3352	249.7	Canadian	–

mega-environment, being significant in at least one macro-environment within the mega-environment. In the Canadian mega-environment, the MQM method detected a QTL that was associated with seed yield at a specific environment (QTL_{SP}) in the interval tagged by Satt-259 and Satt576 on LG O (Table 2). One QTL_{SP} was identified using the same method in the Chinese mega-environment in the interval tagged by Satt194 and SOYGPA, which was located on LG C1 (Table 2). Neither of these markers (Satt194 or 259) have been previously linked to a QTL (SoyBase 2008). No QTL_{SP} were detected between unlinked markers when the single-factor ANOVA method was performed.

The parental line contributing the beneficial allele for both seed yield QTL_{SP} was Heinong #38. Thus, the beneficial allele came from the exotic parent in the Canadian mega-environment and from the adapted parent in the Chinese one (Table 4).

Epistatic effects

Four out of the total of seven seed yield QTL reported (QTL_U and QTL_{SP}) had a significant epistatic interaction with a region located on a different LG (Table 5). Epistasis was only observed among QTL_U but not with QTL_{SP}. An interaction was observed between a seed yield QTL_U tagged by Satt139 and two different regions (linked to Satt002 and Satt288) in both mega-environments. Satt002 is located on LG D2 and was previously reported to tag a QTL associated with seed size (Panthee et al. 2005) and seed yield (Orf et al. 1999a; Reinprecht et al. 2006; Reyna and Sneller 2001). The second marker, Satt288, which is located on LG G, was previously reported as linked to a QTL associated with palmitic, oleic and stearic acid concentrations (Reinprecht et al. 2006).

The strongest seed yield QTL, a QTL_U tagged by Satt100, interacted with two regions, one linked to Satt543 and the other to Satt528 at a specific environment, in the Canadian mega-environment (Table 5). Both interactions were with markers located on LG D2 that mapped closely to each other in this study. Satt543 has been reported as linked to a QTL associated with sclerotinia stem root resistance (Arahana et al. 2001) while Satt528 has not been reported as linked to any trait (SoyBase 2008).

An interaction was observed between a seed yield QTL_U linked to Satt162 and two different unlinked regions tagged by Satt394 and Sat_126 in both mega-environments. Satt394, which is located on LG G has been reported as linked to a QTL associated with sclerotinia stem root resistance (Arahana et al. 2001), seed size (Hyten et al. 2004) and protein content (Reinprecht et al. 2006). Sat_126, which has been linked to a QTL_U (Table 3) in this study, had not been reported previously for any trait (SoyBase 2008).

The amount of phenotypic variation explained by the epistatic interactions (R^2) differed among the QTL_U and ranged between 9 and 14%, with the highest value corresponding to a seed yield QTL_U linked to Satt139 (Table 5).

Discussion

If the goal of a breeding program is to increase yield, the objective could be reached by using the best high-yielding adapted parents. Most often, the best parents will be elite lines that are adapted to the target environment. This is particularly true in soybean breeding, where the gene pool often comprises mostly commercial cultivars (Kabelka et al. 2004). The reason is that results of bi-parental crossing and selection obtained with elite cultivars as parents are often better than those obtained with populations derived from a cross between an elite line and a PI (Smalley et al. 2004). Thompson and Nelson (1998), however, demonstrated that it was possible to select for yield increase in experimental lines derived from crosses where one parental line was a PI. The authors suggested that *Glycine max* accessions could be a possible source of high-yielding alleles for future breeding.

Previous reports of soybean breeding for high yield in which a PI was used did not necessarily involve the use of an exotic parent that was high yielding in its own area of adaptation. Instead, the PI was often an old plant introduction rather than a modern cultivar. Moreover, the mapping population derived from such crosses was not grown in both, the area of adaptation of the adapted parent and the one for the exotic parent. This study was the first one that attempted to use both parents as high yielding and adapted in reciprocate mega-environments and tested them both, the elite parent's adaptation mega-environment (e.g., North America) and the exotic parent's environment (e.g., China). In such way, each parent was considered as either the adapted or exotic parent depending on the testing mega-environment. For example, OAC Millennium was evaluated as an elite line in the Canadian mega-environment and as a PI in the Chinese one. The reverse is true for the Chinese parent, Heinong #38. The concurrent field evaluation of the derived RIL population gave us a unique opportunity to differentiate between two types of QTL that were associated with seed yield: QTL_U and QTL_{SP}. Five seed yield QTL_U were found on four different LG with one on LG C1, I and K and two on LG C2, respectively. On the other hand, seed yield QTL_{SP} was distributed on two LG. Both the adapted parent and the non-adapted one contributed favorable alleles to the RIL population. For example, within the QTL_U identified in this study, the ratio of favorable alleles was 3:2 for alleles coming from the Chinese and Canadian parent, respectively. Such results would not be possible to obtain if the mapping population was only tested in North America. Rather than having a simple marker-assisted selection objective in mind, we attempted to gain a better understanding of how different QTL behaved under different mega-environments that are relevant for high-seed yield allele introgression.

Broad sense heritability for seed yield was estimated for all environments in the period between 2004 and 2006 and ranged from 64% at Harbin, China, in 2004 to 89% at Tavistock, Canada, in 2004. These estimates are comparable to those reported by Guzman et al. (2007) using three backcross-derived populations but higher than those reported for the Essex × Forrest population (Yuan et al. 2002; Kassem et al. 2006).

Table 5 Epistatic effects and amount of phenotypic variation explained by the interacting markers tested across different environments in Canada and China, from 2004 to 2006

Canada						China	
Woodstock 2004	R^2	Tavistock 2004	R^2	Tavistock 2005	R^2	Harbin (1) 2006	R^2
Satt139 × Satt288	14	Satt100 × Satt528	9	Satt139 × Satt002	9	Satt139 × Satt002	9
		Satt100 × Satt543	9	Satt162 × Satt394	12	Sat_126 × Satt162	12

Three QTL_U (linked to Satt100, Satt139 and Satt277) explained most of the seed yield variation observed. Satt277 has been previously reported in different studies as a marker being linked to a QTL associated with yield in soybeans (Orf et al. 1999a, b; Reinprecht et al. 2006; Smalley et al. 2004; Specht et al. 2001). Orf et al. (1999b) study, in particular, reported Satt277 to be linked to a QTL associated with yield in both North and South America where epistatic effects were observed between Satt277 and B172_2, a marker found in LG A2 (Orf et al. 1999b). The results indicated that Satt277 was identified in both mega-environments but no epistatic effects were observed in any macro-environment. The favorable allele came from the PI in the study reported by Orf et al. (1999b). In this study, it came from the Chinese parent, which was the PI in the Canadian mega-environment but the adapted parent in the Chinese one. These results confirmed the seed yield QTL_U identified by Satt277. Since B172_2 was not included in the analysis, the presence or absence of an interaction with Satt277 could not be evaluated. It would be of interest to evaluate, in a future study, whether or not an epistatic effect exists between B172_2 and Satt277, to determine if the epistasis for these regions depends on the genetic background and/or on specific environmental effects (Orf et al. 1999b). In four LG, seed yield QTL (QTL_U and QTL_{SP}) was mapped closer to reported yield QTL. In LG C1, the seed yield QTL_U that was identified by Satt139, mapped less than 3 cM from the reported yield QTL detected by Satt399 (Guzman et al. 2007) and less than 5 cM from Satt294, a marker that detected another yield QTL reported independently in a different study (Yuan et al. 2002). On LG C2, the seed yield QTL_U that was identified by Satt100, mapped less than 1 cM from Satt489 and less than 5 cM from Satt205, markers that identified a seed yield QTL reported by Specht et al. (2001). Satt100 was also closed to other two seed yield QTL identified by Satt134 (Wang et al. 2004) and Satt079 (Mansur et al. 1996). The seed yield QTL_U identified by Satt277 mapped close to three yield QTL identified by Satt205, Satt134 and Satt363 (Kabelka et al. 2004; Specht et al. 2001; Wang et al. 2004). A seed yield QTL_{SP} linked to Satt307 mapped less than 2 cM from a reported yield QTL detected by Satt079 (Mansur et al. 1996). In LG K, the seed yield QTL_{SP} identified by Satt046 mapped less than 10 cM from Satt326 and less than 3 cM from Satt337 and Satt167, all of which were reported for yield (Yuan et al. 2002).

Most of the QTL_U interacted with another genomic region. Three markers linked to seed yield QTL_U interacted with other random markers that were not individually associated with seed yield. Two of the interacting markers had been previously associated with seed weight or seed yield. Interaction between seed yield QTL_U were observed but no interactions were found between a QTL_U and a QTL_{SP} or between any two QTL_{SP}. Thus, since most of the variation observed in seed yield was also accounted by the interaction, both the QTL_U and the interacting region would have to be present in order to take full advantage of their effects in increasing yield (Orf et al. 1999b).

In summary, both specific and universal QTL were found across mega-environments. The non-adapted parent contributed favorable alleles in each mega-environment, which in certain cases explained a large amount of phenotypic variation (R^2). Hence, both major and minor QTL were identified. In the Canadian mega-environment, most of the favorable alleles were contributed by the exotic parent, which demonstrates that favorable alleles could be successfully introduced without a yield penalty due to the lack of mega-environment adaptation of the exotic parent. An improved understanding of which QTL regions are present in one or in both mega-environments as well as their interaction with other genomic regions may facilitate a more efficient introduction of alleles from exotic germplasm without a concomitant decrease in the frequency of favorable alleles for yield. The testing of progeny from crosses involving an exotic parent in each area of adaptation offers a new concept for finding the genetic basis for mega-environment adaptation, which could be evaluated and used in other crop species also. Furthermore, marker-assisted selection could be used to identify and introduce QTL_U and QTL_{SP} as well as the interacting genomic regions into different genetic backgrounds into breeding programs reciprocally between sometimes very remote and different mega-environments. The value of this approach is emphasized especially when trying to target alleles from centers of origin or diversity for major crop species.

Acknowledgments We thank Drs. H. R. Boerma (University of Georgia), S. J. Bowley, K. P. Pauls, G. R. Ablett and D. Falk (University of Guelph) for valuable suggestions on the manuscript. Excellent technical assistance was provided by Wade Montminy, Julia Zilka, Yesenia Salazar, Lin Liao, Chris Grainger, Ron Guillemette, Fernando Pegoraro and technical staff in China. Funding support from the National Science and Research Council of Canada (NSERC), Ontario Ministry of Agriculture, Food and Rural Affairs (OMAFRA) and in-kind support from Pioneer Hi-bred International, a DuPont company is gratefully acknowledged.

References

Arahana VS, Graef GL, Specht JE, Steadman JR, Eskridge KM (2001) Identification of QTLs for resistance to *Sclerotinia sclerotiorum* in soybean. Crop Sci 41:180–188

Bernardo R (2002) Breeding for quantitative traits in plants. Stemma Press, Woodbury, p 369

Bowley S (1999) A hitchhiker's guide to statistics in plant biology. Any Old Subject Books, Guelph, p 250

Braun H, Rajaram S, van Ginkel M (1996) CIMMYT's approach to breeding for wide adaptation. Euphytica 92:175–183

Carter TE Jr, Nelson RL, Sneller CH, Zhanglini C (2004) Genetic diversity in soybean. In: Boerma HR, Specht JE (eds) Soybeans: improvement, production, and uses, 3rd edn. American Society of Agronomy, Inc, Madison, pp 303–416

Concibido VC, La Vallee B, Mclaird P, Pineda N, Meyer J, Hummel L, Yang J, Wu K, Delannay X (2003) Introgression of a quantitative trait locus for yield from glycine soja into commercial soybean cultivars. Theor Appl Genet 106:575–582

Cregan PB, Jarvik T, Bush AL, Shoemaker RC, Lark KG, Kahler AL, Kaya N, VanToai TT, Lohnes DG, Chung L, Specht JE (1999) An integrated genetic linkage map of the soybean genome. Crop Sci 39:1464–1490

Gizlice Z, Carter TE Jr, Burton JW (1994) Genetic base for North American public soybean cultivars released between 1947 and 1988. Crop Sci 34:1143–1151

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

Holland JB (1998) EPISTACY: a SAS program for detecting two-locus epistatic interactions using genetic marker information. J Heredity 89:374–375

Huynh B-L, Wallwork H, Stangoulis JC, Graham RD, Willsmore KL, Oslon S, Mather D (2008) Quantitative trait loci for grain fructan concentration in wheat (*Triticum aestivum* L.). Theor Appl Genet 117:701–709

Hyten DL, Pantalone VR, Sams CE, Saxton AM, Landau-Ellis D, Stefaniak TR, Schmidt ME (2004) Seed quality QTL in a prominent soybean population. Theor Appl Genet 109:552–561

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

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

Lark KG, Chase K, Adler F, Mansur LM, Orf JH (1995) Interactions between quantitative trait loci in soybean in which trait variation at one locus is conditional upon a specific allele at another. Proc Natl Acad Sci USA 92:4656–4660

Mansur LM, Orf JH, Chase K, Jarvik T, Cregan PB, Lark KG (1996) Genetic mapping of agronomic traits using recombinant inbred lines of soybean. Crop Sci 36:1327–1336

Neto ALD, Hashmi R, Schmidt M, Carlson SR, Hartman GL, Li SX, Nelson RL, Diers BW (2007) Mapping and confirmation of a new sudden death syndrome resistance QTL on linkage group D2 from the soybean genotypes PI 567374 and 'Ripley'. Mol Breed 20:53–62

Nyquist WE (1991) Estimation of heritability and prediction of selection response in plant-populations. Crit Rev Plant Sci 10:235–322

OMAFRA (2002) Agronomy guide for field crops—publication 811. http://www.omafra.gov.on.ca/english/crops/pub811/1gddchu.htm#crop. Accessed 11 Sept 2007; verified 21 Oct 2007. Ministry of Agriculture, Food and Rural Affairs, ON, Canada

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

Orf JH, Chase K, Adler FR, Mansur LM, Lark KG (1999b) Genetics of soybean agronomic traits: II. interactions between yield quantitative trait loci in soybean. Crop Sci 39:1652–1657

Panthee DR, Pantalone VR, West DR, Saxton AM, Sams CE (2005) Quantitative trait loci for seed protein and oil concentration, and seed size in soybean. Crop Sci 45:2015–2022

Primomo VS, Poysa V, Ablett GR, Jackson CJ, Gijzen M, Rajcan I (2005) Mapping QTL for individual and total isoflavone content in soybean seeds. Crop Sci 45:2454–2464

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

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

SAS ver. 9.3 (2003) SAS Institute Inc, Cary

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

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

SoyBase (2008) SoyBase and the soybean breeder's toolbox . Knowledge and tools for soybean breeders and researchers. http://soybase.agron.iastate.edu/. Accessed 8 Oct 2006; verified 18 Mar 2008. USDA-Iowa State University, IA, USA

Specht JE, Chase K, Macrander M, Graef GL, Chung J, Markwell JP, Germann M, Orf JH, Lark KG (2001) Soybean response to water: a QTL analysis of drought tolerance. Crop Sci 41:493–509

Thompson JA, Nelson RL (1998) Utilization of diverse germplasm for soybean yield improvement. Crop Sci 38:1362–1368

van Oijen JM (2004) Software for the calculation of genetic linkage maps in experimental populations. JoinMap® 4. Kyazma B.V., Wageningen

van Ooijen JM (2004) Software for the mapping of quantitative trait loci in experimental populations. MapQTL® 5. Kyazma B.V., Wageningen

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

Yuan J, Njiti VN, Meksem K, Iqbal MJ, Triwitayakorn K, Kassem MA, Davis GT, Schmidt ME, Lightfoot DA (2002) Quantitative trait loci in two soybean recombinant inbred line populations segregating for yield and disease resistance. Crop Sci 42:271–277

Zhao JY, Becker HC, Zhang DQ, Zhang YF, Ecke W (2005) Oil content in a European × Chinese rapeseed population: QTL with additive and epistatic effects and their genotype-environment interactions. Crop Sci 45:51–59