

RESEARCH

Context-Specific Marker-Assisted Selection for Improved Grain Yield in Elite Soybean Populations

S. A. Sebastian,* L. G. Streit, P. A. Stephens, J. A. Thompson, B. R. Hedges, M. A. Fabrizius, J. F. Soper, D. H. Schmidt, R. L. Kallem, M. A. Hinds, L. Feng, and J. A. Hoeck

ABSTRACT

Despite the importance of grain yield potential to plant breeders and society in general, it has been difficult to identify grain yield quantitative trait loci (QTL) effective for marker-assisted selection (MAS) across a wide range of genetic and/or environmental contexts. However, as genotyping becomes more cost effective, it might be feasible to use preliminary yield trials to model a target genotype within each context and immediately select the progeny that approach that target genotype in real time. In the present study, elite soybean cultivars with residual heterogeneity were leveraged as populations (the genetic context) to detect yield QTL within a limited set of environments (the environmental context), to model a target genotype, and to select subline haplotypes that comprised the target genotype. The yield potential of the selected subline haplotypes were then compared to their respective mother lines in highly replicated yield trials across multiple environments and years. Statistically significant yield gains of up to 5.8% were confirmed in some of the selected sublines, and two of the improved sublines were released as improved cultivars. This context-specific MAS (CSM) approach might also be applicable to the more typical biparental and backcross populations commonly used in plant breeding programs. Factors that can affect the efficiency and applicability of CSM are discussed.

S.A. Sebastian, D.H. Schmidt, J.F. Soper, L. Feng, M.A. Hinds, and J.A. Hoeck, Pioneer Hi-Bred International, Inc., 7300 NW 62nd Avenue, Johnston, IA 50131; L.G. Streit and R.L. Kallem, 810 Sugar Grove Avenue, Dallas Center, IA 50063; P.A. Stephens, 19741 Illinois Highway 26, Princeton, IL 61356; J.A. Thompson, 9 South Jefferson Street, Mascoutah, IL 62258; B.R. Hedges, 7399 Queen's Line, Chatham, ON N7M5L1, Canada; M.A. Fabrizius, 30263 County Highway 1, Redwood Falls, MN 56283. Received 20 Aug. 2009. *Corresponding author (scott.sebastian@pioneer.com).

Abbreviations: CSM, context-specific marker-assisted selection; $G \times E$, genotype \times environment; yield, grain yield potential per unit land area; MAS, marker-assisted selection; QTL, quantitative trait locus (loci); RIL, recombinant inbred line(s), TPE, target population of environments.

Grain yield potential per unit of land area (herein referred to as yield) is typically the most important trait to both breeders and commercial producers of grain crops. Unfortunately, yield is also the most complex trait to characterize from both a phenotypic and genotypic perspective. Although measured and quantified as a single trait, yield is obviously a complex interaction of many genetic and environmental factors that contribute collectively to the final quantitative measurement. Even within a given field environment, yield measurements are confounded with many sources of nongenetic variation such as variations in seed quality, plot size, soil properties, and disease pressure. This makes it difficult, time consuming, and expensive to identify progeny with the highest yield potential across a sample of environments representative of the target population of environments (TPE) relevant to a given breeding program.

For these reasons, it would be highly desirable to identify genetic markers that are diagnostic of yield potential such that

Published in *Crop Sci.* 50:1196–1206 (2010).

doi: 10.2135/cropsci2009.02.0078

Freely available online through the author-supported open-access option.

Published online 20 Apr. 2010.

© Crop Science Society of America | 5585 Guilford Rd., Madison, WI 53711 USA

All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Permission for printing and for reprinting the material contained herein has been obtained by the publisher.

genetically superior progeny can be identified via marker-assisted selection (MAS) before or during the early stages of field testing. Marker-assisted selection for yield could increase breeding efficiency dramatically by concentrating expensive and time-consuming field testing resources on selections less likely to be artifacts of experimental error.

Yield quantitative trait loci (QTL) are often detected within the context of specific soybean breeding populations and specific environments (Guzman et al., 2007; Orf et al., 1999; Reyna and Sneller, 2001). However, yield QTL in soybean that have been validated across a wide range of genetic and environmental contexts are curiously missing from the literature. Even for specific disease tolerance traits, only a subset of the QTL detected within a given population validate across other populations (Pilet et al., 2001; Robertson-Hoyt et al., 2006). Specific studies and extensive literature reviews confirm this same dilemma in other crops species and for other complex traits (Bernardo, 2008; Holland, 2004, 2007; Podlich et al., 2004; Lubberstedt et al., 2008; Xu and Crouch, 2008).

In a thorough review of molecular markers and selection for complex traits, Bernardo (2008) summarizes the variables that can affect QTL detection and confirmation and concedes that “because estimated QTL effects for traits such as grain yield or plant height have limited transferability across populations, QTL mapping for such traits will likely have to be repeated for each breeding population.” This, in turn, begs the question of whether population-specific yield QTL mapping and MAS would be effective and/or practical (Bernardo, 2008). First, the target genotype would have to be determined separately for each population. Second, the QTL detection experiment would need to sample environments representative of the intended TPE. Third, the sampling of progeny from the mapping population would need to be sufficient to adequately detect and estimate the effects of the major QTL (Beavis, 1994). Considering all of the genetic and nongenetic variables affecting the detection and confirmation of yield QTL, it is no wonder that successful yield MAS is so difficult to demonstrate.

The current study investigates the possibility of a context-specific MAS (CSM) approach for improving grain yield. The term “context-specific” is used herein to distinguish it from “population-specific” and to acknowledge that yield QTL are a function of both population-specific (the genetic context) and environmental-specific (the environmental context) factors. Despite the challenges described above, there are many factors that justify a CSM approach for yield. First of all, yield is typically the most important trait in any breeding program. So the apparent limitations of constructing a customized selection index for each context might be worth the trouble. Second, breeding programs are already set up to measure the yield potential of lines from specific populations across environments typical of a given TPE. Third, the large progeny numbers (Beavis, 1994; Bernardo,

2008) required for effective yield QTL modeling and selection are not necessarily a limitation for well-funded breeding programs. Fourth, genetic marker technology is becoming continuously cheaper and faster with time (Holland, 2004). Considering the expense and error associated with pure phenotypic selection for yield, CSM for yield might actually be the best use of marker resources that a breeder can make.

The current study was designed primarily to answer a simple yet important question: Can favorable yield QTL haplotypes detected within a specific context (a specific soybean population tested at a sample of TPE environments) be useful for MAS of superior-yielding progeny for that TPE?

One way to test CSM for yield in soybean would be selection among recombinant inbred lines (RILs) from a specific biparental or backcross population. Another CSM approach would be selection and comparison of near-isogenic lines that differ at a specific genomic region. A compromise between these two extremes was taken in the current experiments: Recombinant inbred lines were extracted from commercially elite soybean cultivars that retained a small fraction of the genetic heterogeneity present in the original cross from which they were derived. The current study is similar in flavor to the methods of Tuinstra et al. (1997) but unique in both purpose and in many details: First, the ultimate goal of CSM was to identify and confirm transgressive yield segregants as opposed to identifying and confirming QTL for specific traits. Second, the methods required for detecting and confirming genetic gain for yield are more demanding than methods required to detect and confirm genetic gain for more simple traits. Third, in this particular application, CSM leveraged commercially elite cultivars as the base populations for further yield improvement.

CSM within commercially elite cultivars has several logistical and commercially appealing advantages: First it permits detection and MAS of multiple yield QTL within the context of a population that has typically been fixed for yield-confounding traits such as relative maturity, plant height, and disease resistance. Second, the base population would have already been characterized and deemed commercially suitable for a given TPE. Third, if a higher-yielding haplotype is selected from the base population, it can be released immediately as an improved version of the original cultivar.

Previous publications clearly acknowledge the existence of genetic heterogeneity and phenotypic selection for specific traits within cultivars of crop species (Fasoula and Boerma, 2005; Tokatlidis et al., 2004; Gordon and Byth, 1972; Higgs and Russell, 1968). This heterogeneity is mainly the consequence of the fact that the original cultivars were derived from single plants at a relatively early generation of inbreeding. Many commercially elite soybean cultivars, including cultivars used in the current study, are the inbred descendants of a single F_3 or F_4 plant derived from a specific biparental cross. Seed from the selected plant is then multiplied by subsequent generations of self-pollination and seed bulking. The

resulting lines are still F_3 - or F_4 -derived but are typically released as commercial cultivars at a generation of $F_{3:8}$, $F_{4:9}$, or later. These commercial or precommercial lines can therefore be considered as populations with residual heterogeneity at a predictable fraction of the loci that were heterozygous in the original F_1 of the biparental cross. Due to the late stage of inbreeding at the time of commercial release, virtually all of this heterogeneity exists as a mixture of homozygous plants that contain one of the alternate alleles present in the original parents of the cross from which the cultivar was derived.

After the initial single plant selection and during subsequent inbreeding, precommercial soybean lines are often further purified (by phenotypic selection and/or MAS) to be more uniform for disease resistance, maturity, height, and other traits. For example, H43—one of elite cultivars used in the current study—was originally selected as a single F_3 plant. When H43 was at the $F_{3:7}$ generation, a second purification step was taken to make the variety more uniform for visually observable agronomic traits. This was done by pulling approximately 100 individual F_7 plants, planting the resulting F_7 -derived sublines in separate rows in an observation block, and selecting 20 sublines that were more uniform for relative maturity, plant height, and standability than the remainder of the population. So, although H43 is still an F_3 -derived cultivar, it went through a second genetic bottleneck at the F_7 generation. These genetic bottlenecks can significantly skew allele frequencies at some loci that were originally equal (50:50) in the original single plant selection. Regardless of the final ratio of alleles at such heterogeneous loci within the cultivar, genetic gain should be achievable by identification and purification of any yield-favorable haplotypes at said loci.

From this perspective, many elite soybean cultivars can be viewed as heterogeneous mother line populations from which genetically distinct sublines can be extracted. Based on diploid Mendelian theory, F_3 -derived cultivars can be expected to be heterogeneous at an average of 1/4 of the loci that were heterozygous in the original F_1 (Table 1) (Hallauer and Miranda, 1981). For example, if 24 unlinked genomic regions were heterozygous in the F_1 from the original biparental cross, one can expect an average of 6 heterogeneous regions within any given F_3 -derived line from that cross. Of course, specific cultivars can vary quite a bit from the average expectation (Table 1) due to random variations around the average expectation and/or any additional purification practices imposed during the inbreeding process.

MATERIALS AND METHODS

Detection of Heterogeneity Within Each Mother Line

Nine elite soybean mother line populations (Table 2) were chosen for yield QTL detection experiments in 2004. These lines were already being grown commercially or were soon to be released as commercially elite cultivars. They are referred to here as mother lines or mother line populations because they

Table 1. Proportion of loci segregating within and among inbred lines during inbreeding

Generation of inbreeding	Proportion of loci segregating	
	within lines	among lines
F_1	1	0
F_2	1/2	1/2
F_3	1/4	3/4
F_4	1/8	7/8
F_5	1/16	15/16
F_∞	0	1

were treated as heterogeneous populations for the purpose of extracting improved sublines. Mother line names were coded with a unique letter followed by a two digit number indicating relative maturity; for example, one mother line with relative maturity of late group II was coded as “E29.”

Each of the nine mother lines comprised the inbred progeny derived from a single F_3 or F_4 plant from a biparental cross. Therefore, the mother lines could be expected to retain a fraction (Table 1) of the genetic diversity that existed in the original randomly segregating population from which the mother line was selected.

Heterogeneity within each mother line was determined by fingerprinting a bulk sample of leaf tissue DNA from 8 random plants plus individual leaf tissue DNA from 8 additional random plants of each mother line with a set of 100 prioritized genetic markers that were polymorphic within the elite soybean gene pool of Pioneer Hi-Bred International. Based on both the bulk and individual plant samples, heterogeneity was detected at specific marker loci within each mother line population (Table 3). For the purposes of this study, the actual allele numbers listed in Table 3 are irrelevant except for the purpose of detecting and selecting a specific haplotype at a given genetic locus.

The marker prioritization method mentioned above, herein nicknamed “breeding bias,” is described in detail by Sebastian et al. (1995) and has since been described in several other studies in both soybean and corn (Hanafey et al., 1998; Smalley et al., 2004; Feng et al., 2006). Independent of the current study, we used the breeding bias method to scan a larger set of approximately 600 genomic markers for genomic yield QTL “hotspots.” A yield hotspot is defined as a genomic region demonstrating evidence of nonrandom shifts in allele frequency resulting from 50+ years of recurrent selection for yield potential within the elite soybean gene pool adapted to the central U.S. soybean production region. Unpublished data from many internal trials and from the previously cited literature indicated that yield QTL effects (even at genomic hotspots) are notoriously unpredictable in any given context. Therefore, breeding bias was used specifically as a tool for reducing genotyping costs by focusing lab resources on genomic regions with prior evidence of agronomic importance. We were careful not to make assumptions about specific allele effects on yield at said hotspots within the context of any mother line population. Instead, the CSM procedure described below was used to determine the significance, direction, and magnitude of allele effects at said hotspots within the context of each mother line population.

Detection of Yield QTL Within Each Mother Line at a Small Sample of Environments

During the winter of 2004/2005, at winter nurseries in Argentina and Puerto Rico, a small field plot of each mother line was grown

Table 2. Mother line populations and environments sampled for yield quantitative trait loci (QTL) detection.

Mother line population	Zone of adaptation (TPE [†])	TPE environments sampled in 2005 for QTL detection	No. of subline plots at harvest	Yield		
				Mean kg ha ⁻¹	SD kg ha ⁻¹	CV
A06	North-Central United States	Sabin, MN, block 1	72	3322	332	0.10
		Sabin, MN, block 2	72	2856	215	0.08
B27	Central United States	Princeton, IL	72	3830	387	0.10
		Napoleon, OH	72	1054	576	0.55
		Dallas Center, IA	72	3559	519	0.15
C27	Central United States	Princeton, IL	72	3955	413	0.10
		Napoleon, OH	72	1247	438	0.35
		Dallas Center, IA	69	3873	332	0.09
D28	Central United States	Princeton, IL	72	4105	379	0.09
		Napoleon, OH	72	1669	681	0.41
		Dallas Center, IA	72	3484	526	0.15
E29	Central United States	Princeton, IL	72	3943	412	0.10
		Napoleon, OH	72	1628	423	0.26
		Dallas Center, IA	72	3463	486	0.14
F31	Central United States	Princeton, IL	71	4192	344	0.08
		Napoleon, OH	72	2166	564	0.26
		Dallas Center, IA	72	3688	509	0.14
G31	Central United States	Princeton, IL	72	3976	383	0.10
		Napoleon, OH	72	2534	456	0.18
		Dallas Center, IA	72	3980	571	0.14
H43	Midsouthern United States	Mascoutah, IL, block 1	72	5013	433	0.09
		Mascoutah, IL, block 2	72	5441	455	0.08
		Ivesdale, IL	72	3956	390	0.10
I48	Midsouthern United States	Mascoutah, IL	72	5281	340	0.06
		Proctor, AR	72	3053	499	0.16
		Earle, AR	72	3120	559	0.18

[†]TPE, target population of environments.

so that approximately 300 single plants of each mother line could be individually genotyped, allowed to self-pollinate, and harvested to produce an array of RIL sublines. During the growing season, leaf tissue from each plant was sampled, prepared for genotyping, and finally genotyped with the genetic markers previously determined to be heterogeneous within its respective mother line (Table 3). At maturity, the seed from each genotyped plant was then harvested and bulked to comprise a unique subline with a known haplotype at each of the heterogeneous marker loci. The 300 plants selected from each mother line included more than were actually needed for the study. The extra plants were genotyped in case of plant death or in case some plants did not produce enough seed for the subsequent subline yield test in the United States.

A resource-efficient field experiment was then conducted in the United States during the summer of 2005 to (i) measure the yield (phenotype) of each subline within a field environment representative of the TPE relevant to its mother line, (ii) to determine if any of the heterogeneous marker loci were associated with yield differences among sublines from a given mother line population, and (iii) to determine which alleles were yield favorable and potentially useful for MAS within their specific mother line population. Like any QTL analysis, the goal of the subline yield trial was to use the power of allele replication (i.e., data averaging) to mitigate the error associated with yield measurements of individual subline plots such

that yield-favorable alleles could be identified. However, the intent was not to identify generally favorable QTL alleles but to develop a target genotype of favorable QTL alleles that was customized for real-time selection of transgressive segregants from the same population and TPE in which the favorable alleles were detected.

Subline yield trials from a given mother line population were planted in two to three blocks of 72 sublines (72 entries) per block. Most blocks were planted at completely different geographical locations (Table 2), but some blocks were merely placed in adjacent fields of the same farm. The chosen geographical locations were representative of the TPE (Table 2) for which each mother line was adapted. For example, 72 random sublines from mother line population C27 were planted in a single block at a farm in Princeton, IL, another set of 72 random sublines were planted in Napoleon, OH, and a third set of 72 random sublines were planted in Dallas Center, IA (Table 2). In some blocks, individual plots were lost due to attrition from rain gullies or from planting, tillage, or harvesting errors. This is why, for example, only 69 of the 72 original plots of C27 sublines were harvested from the Dallas Center environment (Table 2). At locations where multiple mother lines were being tested, sublines were blocked by mother line and each mother line block was treated as a separate experiment.

Each field plot comprised a single row of a given subline. Each row was 1.5 m long with a planting density of 30 seeds m⁻¹ of row.

Table 3. Genetic heterogeneity detected within each of nine mother line populations.

Marker	Map position		Multiple alleles detected within each of nine mother lines.								
	Chromosome	Position (cM)	A06	B27	C27	D28	E29	F31	G31	H43	I48
S60350TB	A1	46.0			1 & 3				1 & 3		
Satt429	A2	184.0									3 & 4
Satt556	B2	86.1				2 & 4		4 & 5			
Satt190	C1	99.0	1 & 4†								
Satt338	C1	173.0									1 & 6
Satt307	C2	168.3								1 & 4	
Satt216	D1b	8.0					1 & 4				1 & 4
Satt389	D2	93.8						4 & 6			
Satt343	F	1.7	3 & 5	3 & 9					3 & 9		3 & 9
S60227TB	F	84.0		1 & 2							
Satt335	F	126.2		2 & 3							2 & 3
Satt522	F	165.8							4 & 5		
Satt594	G	61.3									4 & 5
Satt352	G	72.4			1 & 4						1 & 4
Satt566	G	72.4			1 & 3						
Satt353	H	8.0								2 & 4	
Satt442	H	42.3				2 & 5				2 & 6	
Satt279	H	77.3								1 & 6	
Satt181	H	125.3			3 & 5						3 & 5
Satt292	I	77.4	2 & 3	2 & 5							
Satt249	J	10.5		3 & 4		1 & 3					
Sag1223	J	19.6	1 & 4								
Sac1699	J	26.4			1 & 3						
Satt406	J	68.0	6 & 9								
Satt431	J	118.0	3 & 5							5 & 6	
Satt712	J	128.9						4 & 9			
Satt242	K	17.9					4 & 5	4 & 5			
Satt544	K	72.8								2 & 3	
Satt398	L	34.6					3 & 4				
Satt497	L	42.3					3 & 5				
S60375TB	L	100.0	2 & 9		2 & 4						
Sag1048	M	84.2			3 & 4				2 & 3		
Satt175	M	91.1							5 & 8		
Sat330DB	M	180.5				1 & 4					
Satt259	O	37.7	2 & 4					4 & 5			
Satt420	O	60.5									1 & 2
Satt477	O	103.8	1 & 3								

†A designation such as '1 & 4' indicates that both alleles 1 and 4 were detected within cultivar A06 with marker Satt190.

Rows were spaced 0.8 m apart from side to side and 1 m apart from end to end. Sublines were randomly assigned to field locations and to rows within locations. In the fall of 2005, the seed from each subline plot was harvested, weighed, and adjusted to 13% moisture. Yield measurements were converted to a kg ha⁻¹ basis.

Yield QTL effects (Table 4) were determined separately within the context of each mother line population and sample of TPE environments using a linear mixed model ANOVA (PROC MIXED; SAS Institute, 2001). The model used was:

$$Y_{ijk} = U + M_i + L_j + ML_{ij} + \epsilon_{ijk}$$

$$L_j \sim iid N(0, \sigma^2_{Location}), ML_{ij} \sim iid N(0, \sigma^2_{ML}), \text{ and } \epsilon_{ijk} \sim iid N(0, \sigma^2_{Residual})$$

where Y_{ijk} = observed plot yield, U = overall mean, M_i = marker QTL effect within a given mother line (fixed), L_j = location effect (random), ML_{ij} = marker × location effect (random), and

ϵ_{ijk} = residual error. In most cases, marker effects within mother lines were considered statistically significant at $P \leq 0.25$.

The logic for relaxing probability values above the typical 0.05 for detection of QTL for traits of low heritability is well explained by Moreau et al. (1998) and Bernardo (2008). In short, Moreau et al. showed via simulation that the consequences of increasing the rate of false positives (Type I errors) is less detrimental than the consequences of false negatives (Type II errors). In the case of the current study, when a Type I error is made for detecting yield QTL (i.e., selection is imposed for a nonsignificant QTL allele), the result is most likely a neutral effect on genetic gain. However, Type II errors represent favorable alleles that could have been used for MAS but were ignored because the statistical cutoff for QTL detection was too stringent.

At statistically significant loci, the allele associated with the highest yield mean was considered the favorable allele for the purpose of selecting higher-yielding sublines within the

Table 4. Homozygous marker genotype yield contrasts within each mother line population.

Mother line population	Marker	Chromosome	Position	Genotype contrast	No. sublines with each genotype	Yield		Yield difference (kg ha ⁻¹)	% Variance explained	Significance of yield difference [P(t)]	Favorable haplotype for CSM†	Notes
						allele1	allele2					
A06	Satt190	C1	99.0	1_1 vs. 4_4	63 vs. 59	3081	3095	-14	0.05	0.88		
A06	Satt343	F	1.7	3_3 vs. 5_5	50 vs. 58	3126	3021	105	3.36	0.06	3_3	
A06	Satt292	I	77.4	2_2 vs. 3_3	44 vs. 77	3123	3067	55	1.13	0.57		
A06	Sag1223	J	19.6	1_1 vs. 4_4	55 vs. 76	3102	3064	38	0.41	0.47		
A06	Satt406	J	68.0	6_6 vs. 9_9	32 vs. 40	3112	2994	118	4.14	0.09	6_6	
A06	Satt431	J	118.0	3_3 vs. 5_5	54 vs. 54	3122	3100	22	0.14	0.70		
A06	S60375TB	L	100.0	2_2 vs. 9_9	59 vs. 69	3066	3123	-57	3.88	0.26		
A06	Satt259	O	37.7	2_2 vs. 4_4	95 vs. 31	3097	3073	24	0.12	0.85		
A06	Satt477	O	103.8	1_1 vs. 3_3	29 vs. 96	3137	3068	70	0.69	0.76		
B27	Satt343	F	1.7	3_3 vs. 9_9	128 vs. 52	2873	2670	203	3.25	0.18	3_3	B27 became obsolete for several re-
B27	S60227TB	F	84.0	1_1 vs. 2_2	60 vs. 127	2722	2845	-122	1.32	0.12	2_2	sons and was not used for CSM
B27	Satt335	F	126.2	2_2 vs. 3_3	82 vs. 123	2810	2824	-14	0.02	0.86		
B27	Satt292	I	77.4	2_2 vs. 5_5	120 vs. 85	2866	2752	113	2.00	0.11	2_2	
B27	Satt249	J	10.5	3_3 vs. 4_4	121 vs. 81	2784	2862	-78	0.57	0.29		
C27	S60350TB	A1	46.0	1_1 vs. 3_3	114 vs. 55	3062	2890	173	4.32	0.01	1_1	
C27	Satt352	G	72.4	1_1 vs. 4_4	111 vs. 64	3013	3092	-79	0.94	0.20		
C27	Satt566	G	72.4	1_1 vs. 3_3	108 vs. 61	2987	3096	-108	1.73	0.09	3_3	
C27	Satt181	H	125.3	3_3 vs. 5_5	19 vs. 161	3211	3016	195	2.34	0.04	3_3	
C27	Sac1699	J	26.4	1_1 vs. 3_3	64 vs. 89	2949	3098	-149	3.46	0.18	3_3	
C27	S60375TB	L	100.0	2_2 vs. 4_4	51 vs. 141	3063	3013	50	0.30	0.45		
C27	Sag1048	M	84.2	3_3 vs. 4_4	61 vs. 136	2940	3067	-126	2.10	0.04	4_4	
D28	Satt556	B2	86.1	2_2 vs. 4_4	88 vs. 96	3170	3028	142	2.23	0.08	2_2	D28 was not used for CSM due to limited
D28	Satt442	H	42.3	2_2 vs. 5_5	116 vs. 73	3124	3046	77	1.08	0.34		QTL detection with the markers used
D28	Satt249	J	10.5	1_1 vs. 3_3	109 vs. 83	3112	3043	68	0.40	0.49		
D28	Satt330DB	M	180.5	1_1 vs. 4_4	87 vs. 63	3093	3063	30	2.73	0.74		
E29	Satt216	D1b	8.0	1_1 vs. 4_4	137 vs. 71	3034	2963	71	0.65	0.62		E29 was not used for CSM due to limited
E29	Satt242	K	17.9	4_4 vs. 5_5	41 vs. 149	2961	3013	-52	0.24	0.50		QTL detection with the markers used
E29	Satt398	L	34.6	3_3 vs. 4_4	98 vs. 95	2914	3110	-196	4.92	0.13		
E29	Satt497	L	42.3	3_3 vs. 5_5	87 vs. 76	3107	2870	237	7.08	0.00	3_3	
F31	Satt556	B2	86.1	4_4 vs. 5_5	161 vs. 39	3370	3326	45	0.19	0.84		F31 was not used for CSM due to no
F31	Satt389	D2	93.8	4_4 vs. 6_6	131 vs. 28	3340	3279	61	0.09	0.86		detection of QTL with the markers used
F31	Satt712	J	128.9	4_4 vs. 9_9	142 vs. 54	3367	3284	83	0.67	0.49		
F31	Satt242	K	17.9	4_4 vs. 5_5	54 vs. 129	3329	3330	-1	0.00	0.99		
F31	Satt259	O	37.7	4_4 vs. 5_5	38 vs. 138	3341	3404	-63	0.30	0.64		
G31	S60350TB	A1	46.0	1_1 vs. 3_3	104 vs. 93	3555	3431	124	1.65	0.26	1_1†	S60359TB and Satt343 were used for
G31	Satt343	F	1.7	3_3 vs. 9_9	85 vs. 94	3568	3423	146	2.36	0.38	3_3†	CSM despite their marginal significance;
G31	Satt522	F	165.8	4_4 vs. 5_5	29 vs. 102	3526	3518	7	1.82	0.94		explained in text.
G31	Sag1048	M	84.2	2_2 vs. 3_3	54 vs. 135	3613	3438	175	2.73	0.02	2_2	
G31	Satt175	M	91.1	5_5 vs. 8_8	90 vs. 104	3429	3555	-125	1.77	0.44		

(cont'd)

Table 4. Continued.

Mother line population	Marker	Chromosome	Position	Genotype contrast	No. sublines with each genotype	Yield allele1 (kg ha ⁻¹)	Yield allele2 (kg ha ⁻¹)	Yield difference	% Variance explained	Significance of yield difference [P(t)]	Favorable haplotype for CSM†	Notes
H43	Satt307	C2	168.3	1_1 vs. 4_4	89 vs. 80	4709	4897	-187	4.04	0.22	4_4	
H43	Satt353	H	8.0	2_2 vs. 4_4	72 vs. 101	4811	4789	22	0.08	0.72		
H43	Satt442	H	42.3	2_2 vs. 6_6	33 vs. 140	4960	4782	177	2.71	0.32		
H43	Satt279	H	77.3	1_1 vs. 6_6	24 vs. 164	5056	4763	293	5.26	0.00	1_1	
H43	Satt431	J	118.0	5_5 vs. 6_6	128 vs. 44	4711	5018	-307	10.56	0.15	6_6	
H43	Satt544	K	72.8	2_2 vs. 3_3	168 vs. 21	4754	5062	-308	5.14	0.00	3_3	
I48	Satt429	A2	184.0	3_3 vs. 4_4	93 vs. 89	3930	3694	236	6.31	0.00	3_3	Satt216 was used for CSM despite statistical insignificance; explained in text.
I48	Satt338	C1	173.0	1_1 vs. 6_6	114 vs. 78	3832	3816	16	0.05	0.90		
I48	Satt216	D1b	8.0	1_1 vs. 4_4	29 vs. 90	3669	3802	-133	2.35	0.52	4_4‡	
I48	Satt343	F	1.7	3_3 vs. 9_9	119 vs. 70	3811	3822	-11	0.01	0.88		
I48	Satt335	F	126.2	2_2 vs. 3_3	85 vs. 112	3803	3800	3	0.01	0.98		
I48	Satt594	G	61.3	4_4 vs. 5_5	153 vs. 41	3854	3715	139	1.27	0.33		
I48	Satt352	G	72.4	1_1 vs. 4_4	33 vs. 103	3756	3840	-84	3.22	0.38		
I48	Satt181	H	125.3	3_3 vs. 5_5	58 vs. 118	3847	3769	78	0.54	0.45		
I48	Satt420	O	60.5	1_1 vs. 2_2	36 vs. 160	4020	3801	219	4.54	0.23	1_1	

†CSM, context-specific marker-assisted selection

‡These markers were used for marker-assisted selection despite their insignificance at $P(t) < = 0.25$. The reason is explained in the text.

context of a given mother line population (Table 4). Exceptions to the 0.25 cutoff for statistical significance are explained below and are also noted in Table 4.

Genotypic Selection of Putatively Improved Sublines

Although E29 showed significance at 2 markers (Satt398 and Satt497), these markers were linked within 8 cM and therefore considered diagnostic of only one yield QTL region. Likewise, only one significant yield QTL region was detected within mother line D28. Mother lines E29 and D28 were therefore not pursued for CSM because they showed limited potential for genetic gain with the marker coverage available at the time of the study. Mother line B27 showed evidence of multiple significant QTL but fell out of favor due to inferior agronomic performance in relation to other precommercial cultivars tested during 2005. For this reason, B27 was not pursued for improvement via CSM. Out of the original nine mother line populations used for yield QTL detection, five (A06, C27, G31, H43, and I48) were still considered commercially viable by the end of 2005 and also showed evidence of multiple yield QTL regions that could be leveraged for CSM and possible genetic gain.

After establishing the target genotype for selection within each of the five mother line populations noted above, the next step was to select those sublines that had the complete complement of significantly favorable alleles detected in the yield QTL analysis (Table 4). For example, 11 out of the 216 sublines of H43 tested in short rows in 2005 were homozygous for the favorable allele at Satt307, Satt279, Satt431, and Satt544. In cases where closely linked markers showed significant effects on yield (such as Satt352 and Satt566 within mother line C27), the marker with the best statistical significance (Satt566 in this case) was used for selection purposes. The selected sublines were those with the complete set of favorable alleles shown in the favorable haplotype column of Table 4.

Exceptions to the 0.25 cutoff for statistical significance of S60359TB ($P = 0.26$) and Satt343 ($P = 0.38$) within mother line population G31 and Satt216 ($P = 0.52$) within mother line population I48 (Table 4) were the result of a statistical error that was caught during the review process. Specifically, the authors originally incorrectly used residual error variance (ϵ_{ijk}) to test the significance of marker effects. However, marker \times location interaction variance (ML_{ij}) contributes to the estimated variation among marker main effects, thus marker \times location variation contributes to the appropriate error term for testing marker effects. Fortunately, the inclusion of ML_{ij} in the denominator of F tests of marker effects (reflected in the QTL Probability(t) [$P(t)$] values in Table 4) did not significantly affect the target genotypes used for selection purposes. The only difference is that some potentially nonsignificant alleles were included in the target genotypes for selection of sublines from G31 and I48 along with the other alleles that were statistically significant at $P \leq 0.25$. The effect of including nonsignificant alleles in the target genotype was most probably neutral because the appropriate $P(t)$ value indicates, at worst, that neither allele was favorable.

Although not an issue in the current examples, if none of the progeny contained all of the favorable alleles detected in the QTL analysis, one could simply prioritize the alleles based on their estimated effects and select those progeny that had as

Table 5. Subline bulks advanced into replicated field trials in 2006 and 2007.

Mother line	No. sublines tested in 2005	No. sublines selected for improved bulk	Name of selected haplotype bulk
A06	195	12	ZB06M06
C27	214	11	ZB27L06
G31	217	12	ZB31T06
H43	201	11	ZB43F06
I48	217	13	ZB48W06

many of the favorable alleles as possible (Bonnett et al., 2005). For studies where genome-wide marker saturation is available, many QTL are detected, and many nonadditive interactions are anticipated, one could use genome-wide modeling methods to sort progeny based on their unique genotypes (Meuwissen et al., 2001; Bernardo and Yu, 2007).

Bulking of Sublines with the Putatively Favorable Haplotype

Approximately 0.5 kg of seed from each subline was available from the short row field test grown and harvested in 2005. Hence the 2005 field test served the purpose of both QTL analysis and as a seed source for subsequent replicated testing of selected sublines. Equal quantities of seed from multiple sublines comprising the favorable haplotype detected in each mother line population were pooled to create a selected haplotype bulk from each mother line (Table 5).

There are both genetic and logistical reasons that a bulk of sublines with the favorable haplotype was used to confirm genetic gain over the mother line as opposed to comparing individual sublines to the mother line. From a genetic perspective, bulking of multiple subline selections was done to retain as much heterogeneity as possible at non-target loci so that any genetic gain realized by selection could be attributed to the target genotype as opposed to genetic drift (sampling error) at other potentially heterogeneous loci. Non-target loci include those known to be heterogeneous via markers (yet statistically insignificant) and other loci of unknown heterogeneity due to the limited marker coverage available at the time this study was initiated.

Logistically, the bulked subline versus original mother line comparison made it feasible to include them as only two additional entries along with many other precommercial lines of similar maturity in multi-location, multi-year Pioneer Hi-Bred soybean departmental trials. The bulking method therefore minimized the field resources needed in the multi-environment confirmation phase by concentrating testing resources on the comparison of most interest to the study: the CSM haplotype versus the unselected mother line bulk. The trade-off of this design was that the experiment could not simultaneously prove that CSM-selected sublines performed better than phenotypically selected sublines. But this was not the goal of the current study since we were already very aware from decades of experience that selections based on individual progeny row yield phenotypes had very low repeatability (low heritability) in subsequent trials. In fact, the statistical imprecision of individual yield measurements was the key motivation to determine if CSM for yield was even possible. Once CSM was demon-

strated, other studies (still in progress) were initiated to quantify the relative efficiency of CSM versus phenotypic selection.

Confirmation of Genetic Gain across a Broad Sample of Environments

To confirm genetic gain of the selected haplotypes, each selected subline bulk was compared to its respective mother line in highly replicated field trials across many environments and across 2 yr (2006 and 2007). The actual field environments chosen for confirmation of genetic gain were considered to be representative of the TPE for which the mother line was specifically adapted for commercial production (Table 2). Each experimental unit (yield test plot) comprised a single soybean line planted in two rows 3.8 m long and spaced 0.8 m apart. Planting density was 30 seeds m⁻¹ of row. Plots were randomized within complete blocks containing 15 to 40 entries including the mother line, its corresponding selected subline bulk, and other soybean lines also being evaluated for commercial potential. The number of environments and replications per environment varied for each subline to mother line paired contrast. In addition, average yield potential and phenotypic range varied quite a bit from one environment to another. Therefore, subline versus mother line yield contrasts were made with varying levels of replication and statistical precision (Table 6). For example, improved subline bulk ZB43F06 was compared to mother line H43 at 44 different environments (unique fields) representative of the geographic regions where H43 is adapted and commercially grown. Some of these environments had 2 to 3 blocks (replications) of the same contrast; hence, the total number of replications for each contrast was much greater than the number of environments sampled. In total, ZB43F06 was compared to its mother line 106 times across the 44 environments and 2 yr. Grain yield means and statistical significance values (Table 6) were adjusted to remove environment and block-within-environment effects.

It is well known that grain yield differences between soybean lines can be influenced by their relative maturity date. In fact, the potential for confounding effects of relative maturity date on yield potential is one of the reasons that CSM was tested within the context of elite populations that were already very homogeneous in terms of their relative maturity date. However, it is possible that selection for yield QTL could cause a slight difference in relative maturity date between the selected haplotypes and the original mother line. If so, we wanted to ensure that any yield differences detected were not simply the result of selection for QTL affecting maturity date. The general tendency is a positive correlation between maturity date and yield simply because late-maturing soybean lines have more time to grow and produce grain than early-maturing soybean lines. However, specific conditions in any given environment, geographic region, or year can change the direction and magnitude of this maturity effect on grain yield. For example, a late-season drought or early frost in a given geographic region might actually cause a negative association between late maturity and yield. Hence, the relative maturity date of all soybean lines (including mother line and subline) within each of the replicated yield trials was noted and used to determine the average effect of maturity date on yield within the environments sampled. Observed maturity differences, their statistical significance, and the average effect of maturity on yield (considered significant at $R^2 > 0.10$) are reported in Table 6.

Table 6. Yield of subline vs. mother line across multiple environments in 2006 and 2007.

Subline	Mother	Total no. environments	Total no. replications	Subline yield (kg ha ⁻¹)	Mother yield (kg ha ⁻¹)	Yield difference (%)	Significance of yield difference <i>P</i> (<i>t</i>)	Maturity difference (d)	Significance of maturity difference <i>P</i> (<i>t</i>)	Effect of maturity on yield (% per d)
ZB06M06	A06	29	81	3195	3185	0.3	0.85	-0.3	0.43	1.2
ZB27L06	C27	45	89	3843	3698	3.9	0.00	-0.1	0.64	ns [†]
ZB31T06	G31	45	89	3851	3729	3.3	0.01	0.6	0.04	ns
ZB43F06	H43	44	106	3538	3343	5.8	0.00	1.0	0.07	ns
ZB48W06	I48	35	88	3425	3355	2.1	0.21	1.0	0.05	1.2

[†]ns, not significant.

RESULTS

In multi-year multi-environment trials (Table 6), three of the five selected haplotypes (ZB27L06, ZB31T06, and ZB43F06) were significantly higher yielding than their respective mother lines from both a statistical and commercially relevant perspective. The other two subline bulks (ZB06M06 and ZB48W06) were also higher yielding than their respective mother lines but the yield differences were not statistically significant. The following yield differences were observed between the selected subline bulks and their respective mother lines:

Subline bulk ZB43F06 averaged 5.8% higher grain yield ($P = 0.0004$) compared to its mother line H43 across a total of 44 different environments and 106 replications (Table 6). Although ZB43F06 was an average of 1.0 d later than H43 in relative maturity, there was no significant correlation between maturity and yield in the collective set of field experiments used to compare these two lines. Subline bulk ZB27L06 averaged 3.9% higher yield ($P = 0.0000$) than its mother line C27 across a total of 45 different environments and 89 replications. No significant difference in relative maturity was observed between ZB27L06 and its mother line. Subline bulk ZB31T06 averaged 3.3% higher yield ($P = 0.009$) than its mother line G31 across a total of 45 different environments and 89 replications. Although the selected subline was slightly later in maturity (0.6 d) than G31, there was no significant correlation between maturity and yield in this set of field experiments.

ZB48W06 was 2.1% higher yielding than mother line I48 across 35 environments and 88 replications but this was not statistically significant ($P = 0.21$). In addition, ZB48W06 was 1.0 d later in maturity than I48. In these trials, a 1 d maturity difference could explain about 1/2 of the yield difference detected (Table 6). ZB06M06 was only 0.3% higher yielding than mother line A06 in the 29 environments and 81 replications tested ($P = 0.85$); this was a nonsignificant difference.

In summary, out of the nine original mother line populations tested, five showed evidence of multiple yield QTL for genetic gain via CSM and further commercial potential based on the continued acceptability of the mother line per se in 2005. Since we did not attempt CSM within the other four populations, we cannot comment on whether

genetic gain would have been realized via CSM. However, out of the five populations where CSM was attempted, three attempts resulted in a statistically significant yield gain versus the unselected mother line population across a wide range of environmental conditions. Given the bulking process taken to control genetic drift and accounting for possible differences due to relative maturity, the genetic gain observed in the three subline bulks can be attributed to selection of the favorable haplotypes detected in the original yield QTL analyses (Table 4). In addition to demonstrating significant yield gains over their respective mother lines, ZB43F06 and ZB27L06 were released as new commercial cultivars based on their superiority to their mother lines and to other (unrelated) commercial and precommercial lines being tested in the same environments. Although ZB31T06 was significantly higher yielding than its mother line, it was not released as a commercial cultivar because other higher-yielding (but unrelated) lines were available for commercial release in the same TPE.

DISCUSSION

Although the importance of context specificity has been mentioned in many QTL related publications, the concept of CSM has been considered impractical for reasons discussed in the introduction of this paper. But given the ever-decreasing cost of whole-genome genotyping, the practicality of CSM for grain yield (the quintessential trait of interest) needs to be reconsidered. Breeders are already aware that individual yield measurements have very low heritability due to the many sources of experimental error inherent in yield testing. This error is highest during the first year of yield testing, where the replication and precision for measuring each progeny's yield potential is lowest. However, during this same phase of testing, the number of progeny tested (allele replication) and environments sampled can be as high as the breeder needs for an accurate yield QTL analysis of a given context. The most compelling incentive for CSM is to improve the heritability of selections that will be advanced into the resource-intensive confirmation trials that must follow to ensure that a new cultivar will perform well across a wide range of environments. Effective genotypic selection can dramatically improve breeding efficiency by focusing these resources on progeny selections

that are more likely to be true transgressive segregants and less likely to be artifacts of experimental error.

Like any MAS procedure, CSM uses molecular markers as genetic covariates to mitigate the confounding effects of experimental error that reduce the heritability of individual phenotypic measurements. The unique aspect of CSM is that it focuses the power of genetic markers to construct a target genotype customized for a specific population and TPE. This eliminates the requirement to validate QTL across other populations and other environments that lie outside of the TPE. The only validation required for CSM is the confirmation of significant genetic gain of the selected haplotype within the TPE.

It is noteworthy that the current studies required minimal field and marker resources to demonstrate CSM and to release significantly improved commercial cultivars. For example, during the QTL detection phase, a small sample of one to three distinct environments sampled from the larger reference TPE within 1 yr (2005) were needed to identify potentially useful yield QTL within any given genetic context (Table 2). Although one to three environments in 1 yr might appear to be poor sampling of the TPE, this is actually representative of the way commercial soybean breeding programs conduct early-generation yield testing: inbred lines derived from a given population are typically tested in small plots at a single environment that hopefully will be representative of the TPE. But, if the early-generation test environment is not representative of the TPE, this might not be predictive of genotypes that are favorable across the broader sample of TPE environments encountered in subsequent replicated trials (Bernardo, 2008). This could be the explanation why CSM did not result in significant genetic gain within mother lines A06 and I48 even though putatively favorable alleles were identified in the QTL detection phase.

Another factor that can affect the ability to detect yield-favorable alleles is the quality of the yield data (i.e., error variance) within the environments being sampled to detect yield QTL. Simple statistics such as mean, standard deviation, and CV for yield can be used to indicate the relative quality of data derived from different field environments for the purpose of QTL detection. Although differences in mean yield and CV varied quite a bit at QTL detection locations (Table 2), data from all yield trial locations was used to sample the largest number of progeny and environments available. Perhaps environments with high error variance or environments suspected to be unrepresentative of the TPE should be excluded from the QTL analysis so that more valid QTL estimates can be obtained to construct the favorable haplotype for CSM. Breeders prefer testing environments that permit expression of high yield potential yet have low spatial variation in soil type, soil depth, slope, and drainage properties. Such environments are more likely to expose differences in genetic potential and minimize differ-

ences due to nongenetic factors. It seems logical that these environments also should be favored for effective CSM.

As demonstrated in these sublining experiments, approximately 216 small progeny plots were needed to detect QTL and give positive results in some of the F_3 - or F_4 -derived mother line populations. However, the progeny sample size required to accurately estimate QTL effects is clearly a function of how much genetic diversity is being sampled within a given genetic context. In more diverse populations, more haplotype combinations are possible and more progeny are required to sample the total genetic space of the population (Beavis, 1994).

It is likely that more genetic gain could have been realized with better genome marker coverage during the QTL detection and MAS phase of this study. Although it seems logical to assume that the focus of marker resources on genomic hotspots might have reduced the need for more complete marker coverage, this assumption was not tested in the current study. To prove or disprove this assumption, full genome coverage and additional studies are required to compare the “hit rate” of hotspots versus random loci for the detection of significant yield QTL and the allele that is favorable in any given context.

Despite the above considerations to improve CSM, the current experiments do demonstrate that MAS for improved grain yield is possible if focused within a specific genetic and environmental context. Other studies (in progress) are being conducted to quantify the relative efficiency of CSM versus phenotypic selection for yield and to determine the feasibility of CSM within populations of broader genetic diversity such as biparental and backcross populations. However, based on the examples shown here and progress in ongoing experiments, CSM has already been adopted as a major component of MAS strategies known commercially as Accelerated Yield Technology (AYT) at Pioneer Hi-Bred International.

References

- Beavis, W.D. 1994. The power and deceit of QTL experiments: Lessons from comparative QTL studies. p. 250–266. *In Proc. Corn Sorghum Ind. Res. Conf.*, 7–8 Dec. 1994, Chicago, IL. Am. Seed Trade Assoc., Washington, DC.
- Bernardo, R. 2008. Molecular markers and selection for complex traits in plants: Learning from the last 20 years. *Crop Sci.* 48:1649–1664.
- Bernardo, R., and J. Yu. 2007. Prospects for genomewide selection for quantitative traits in maize. *Crop Sci.* 47:1082–1090.
- Bonnett, D.G., G.J. Rebetzke, and W. Spielmeyer. 2005. Strategies for efficient implementation of molecular markers in wheat breeding. *Mol. Breed.* 15:75–85.
- Fasoula, V.A., and H.R. Boerma. 2005. Divergent selection at ultralow plant density for seed protein and oil content within soybean cultivars. *Field Crops Res.* 91:217–229.
- Feng, L., S. Sebastian, S. Smith, and M. Cooper. 2006. Temporal trends in SSR allele frequencies associated with long term selection of yield of maize. *Maydica* 51:293–300.
- Gordon, I.L., and D.E. Byth. 1972. Comparisons among strains of the tobacco cultivar Hicks illustrating variability within a single cultivar. *Queensland J. Agric. Anim. Sci.* 29:255–264.

- Guzman, P.S., B.W. Diers, D.J. Neece, S.K. St. Martin, A.R. LeRoy, C.R. Grau, T.J. Hughes, and R.L. Nelson. 2007. QTL associated with yield in three backcross-derived populations of soybean. *Crop Sci.* 47:111–122.
- Hallauer, A.R., and J.B. Miranda. 1981. p. 32. Quantitative genetics in maize breeding. Iowa State Univ. Press, Ames, IA.
- Hanafey, M.K., S.A. Sebastian, and S.V. Tingey. 1998. Method to identify genetic markers that are linked to agronomically important genes. U.S. Patent 5 746 023. Date issued: 5 May.
- Higgs, R.L., and W.A. Russell. 1968. Genetic variation in quantitative characters in maize inbred lines, I. Variation among and within Corn Belt seed sources of six inbreds. *Crop Sci.* 8:345–348.
- Holland, J.B. 2004. Implementation of molecular markers for quantitative traits in breeding programs—Challenges and opportunities. In T. Fischer (ed.) *New directions for a diverse planet*. Proc. Int. Crop Sci. Congress, 4th, Brisbane, QLD, Australia. 26 Sept.–1 Oct. 2004. Regional Inst. Ltd., Gosford, NSW, Australia.
- Holland, J.B. 2007. Genetic architecture of complex traits in plants. *Curr. Opin. Plant Biol.* 10:156–161.
- Lubberstedt, T., A.E. Melchinger, S. Fahr, and D. Klein. 2008. QTL mapping in testcrosses of flint lines of maize: III. Comparison across populations for forage traits. *Crop Sci.* 38:1278–1289.
- Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829.
- Moreau, L., A. Charcosset, F. Hospital, and A. Gallais. 1998. Marker-assisted selection efficiency in populations of finite size. *Genetics* 148:1353–1365.
- Orf, J.H., K. Chase, T. Jarvik, L.M. Mansur, P.B. Cregan, F.R. Adler, and K.G. Lark. 1999. Genetics of soybean agronomic traits I. Comparison of three related recombinant inbred populations. *Crop Sci.* 39:1642–1651.
- Pilet, M.L., G. Duplan, M. Archipiano, P. Barret, C. Baron, R. Horvais, X. Tanguy, M.O. Lucas, M. Renard, and R. Delourme. 2001. Stability of QTL for field resistance to blackleg across two genetic backgrounds in oilseed rape. *Crop Sci.* 41:197–205.
- Podlich, D. W., C. R. Winkler, and M. Cooper. 2004. Mapping as you go: An effective approach for marker-assisted selection of complex traits. *Crop Sci.* 44:1560–1571.
- Reyna, N., and C.H. Sneller. 2001. Evaluation of marker-assisted introgression of yield QTL alleles into adapted soybean. *Crop Sci.* 41:1317–1321.
- Robertson-Hoyt, L.A., M.P. Jines, P.J. Balint-Kurti, C.E. Kleinschmidt, D.G. White, G.A. Payne, C.M. Maragos, T.L. Molnár, and J.B. Holland. 2006. QTL mapping for Fusarium ear rot and Fumonisin contamination resistance in two maize populations. *Crop Sci* 46:1734–1743.
- SAS Institute. 2001. SAS/STAT user's guide, v 9.1. SAS Institute Inc, Cary, NC.
- Sebastian, S.A., S.V. Tingey, and M.K. Hanafey. 1995. Method to identify genetic markers that are linked to agronomically important genes. U.S. Patent 5 437 697. Date issued: 1 August.
- Smalley, M.D., W.R. Fehr, S.R. Cianzio, F. Han, S.A. Sebastian, and L.G. Streit. 2004. Quantitative trait loci for soybean seed yield in elite and plant introduction germplasm. *Crop Sci.* 44:436–442.
- Tokatlidis, I.S., J.T. Tsialtas, I.N. Xynias, E. Tamoutsidis, and M. Irakli. 2004. Variation within a bread wheat cultivar for grain yield, protein content, carbon isotope discrimination and ash content. *Field Crops Res.* 86:33–42.
- Tuinstra, M.R., G. Ejeta, and P.B. Goldsbrough. 1997. Heterogeneous inbred family (HIF) analysis: A method for developing near-isogenic lines that differ at quantitative trait loci. *Theor. Appl. Genet.* 95:1005–1011.
- Xu, Y., and J.H. Crouch. 2008. Marker-assisted selection in plant breeding: From publications to practice. *Crop Sci.* 48:391–407.