Theor Appl Genet (2011) 122:581–594
DOI 10.1007/s00122-010-1471-x

ORIGINAL PAPER

Mapping quantitative trait loci for seed size traits in soybean (*Glycine max* L. Merr.)

Yu Xu · He-Nan Li · Guang-Jun Li · Xia Wang · Li-Guo Cheng · Yuan-Ming Zhang

Received: 24 May 2010 / Accepted: 11 October 2010 / Published online: 28 October 2010

Abstract Seed size traits in soybean—length, width and thickness—and their corresponding ratios—length-to-width, length-to-thickness and width-to-thickness—play a crucial role in determining seed appearance, quality and yield. In this study, an attempt was made to detect quantitative trait loci (QTL) for the aforementioned seed size traits in $F_{2:3}$, $F_{2:4}$ and $F_{2:5}$ populations from the direct and reciprocal crosses of Lishuizhongzihuang with Nannong 493-1, using multi-QTL joint analysis (MJA) along with composite interval mapping (CIM). A total of 121 main-effect QTL (M-QTL), six environmental effects, eight environment-by-QTL interactions, five cytoplasmic effects and 92 cytoplasm-by-QTL interactions were detected. Fifty-two common M-QTL across MJA and CIM, 21 common M-QTL in more than two populations and 5 M-QTL in all three populations showed the stability of the results. Five M-QTL had higher heritability, greater than 20%. In addition, 28 cytoplasm-by-QTL and 4 environment-by-QTL interactions were confirmed by CIM. Most M-QTL were clustered in eight chromosomal regions. Our results provide a good foundation for fine mapping, cloning and designed molecular breeding of favorable genes related to soybean seed size traits.

Communicated by I. Rajcan.

Y. Xu and H.-N. Li contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-010-1471-x) contains supplementary material, which is available to authorized users.

Y. Xu · H.-N. Li · G.-J. Li · X. Wang · L.-G. Cheng · Y.-M. Zhang (✉)
Section on Statistical Genomics, State Key Laboratory of Crop Genetics and Germplasm Enhancement, College of Agriculture, Nanjing Agricultural University, 1 Weigang Road, Nanjing 210095, China
e-mail: soyzhang@njau.edu.cn

Introduction

The aims of soybean breeding are to increase seed yield and to improve seed quality. Seed yield per unit area is the product of number of plants per unit area, number of seeds per plant and 100-seed weight. The 100-seed weight is affected by seed size, measured by length, width and thickness. Seed size is a major target of breeding, not only as a component of seed yield (Liang et al. 2005) but also as a morphological quality trait (Wilson 1995). In addition, commercial value in international trade partly depends on seed appearance (Cui and Xuan 2007). Therefore, the great economic importance associated with seed size necessitates in-depth study of its genetic basis and developmental mechanism to better understand biological development processes and to facilitate breeding in soybean.

During the past decade, many attempts have been made to dissect the genetic mechanism of seed size in rice, peanut, maize and barley (Zheng et al. 1985; Ayoub et al. 2002; Salas et al. 2006; Rabiei et al. 2004), and the most progress has been made in rice. Grain length in rice is a quantitative trait controlled by multiple major genes (Kazuyoshi and Ayumi 1980) or by polygenes with an additive effect (Shi and Shen 1994). Recently, a molecular quantitative genetics approach has been used to detect quantitative trait loci (QTL) for grain size traits (Lin et al. 1995; Rabiei et al. 2004; Wan et al. 2008). Many QTL for grain weight and size in rice have been detected (Sakamoto and Matsuoka 2008), including the following three that were recently identified: *GS3* (Fan et al. 2006), *GW2* (Song et al. 2007) and *qSW5* (Shomura et al. 2008; Wan et al. 2008). Although the functions of the genes

in these three QTL have not yet been clarified, cloning of such genes will provide the opportunity to characterize the regulatory mechanisms of grain development and create a potential tool for improving grain size.

Although there has been substantial research on the classic inheritance of seed size and seed weight in soybean (Kim et al. 2000; Johnson et al. 2001; Orf et al. 1999; Mansur et al. 1993, 1996; Leroy et al. 1991), few studies have been done on the molecular inheritance of seed size and shape traits. Nelson and Wang (1989) showed that seed size and seed shape can be inherited stably; Cober et al. (1997) indicated that seed shape is independent of seed size, and a similar result has been recently shown in wheat by Gegas et al. (2010). The estimate of heritability is 59–79% for seed shape and 19–56% for seed size (Cober et al. 1997). Liang et al. (2005) carried out an incomplete diallele cross of eight varieties with their F_1 and F_2 populations and showed that the inheritance of seed weight and seed length (SL) is mainly controlled by cytoplasmic effects, whereas seed width (SW) and thickness (ST) are mainly controlled by maternal effects. Within the above studies, the collective properties of genes were studied, and the position and the effects of single QTL or gene are not clear. Therefore, a method of locating QTL should be carried out (Mansur et al. 1993). Recently, Salas et al. (2006) mapped the QTL responsible for seed shape traits with three densely mapped recombinant inbred populations. A total of 19 significant QTL in 10 linkage groups was detected for all seed shape traits: only one QTL was stable across populations and environments, and six were stable in at least two populations in both environments. Three QTL for SL were detected in the linkage groups F, J and M by Li et al. (2008). In addition, seven QTL for SL, three for SW and three for ST were mapped by Liang et al. (2008). However, environment-by-QTL and cytoplasm-by-QTL interactions have not been reported.

The objectives of this study were to obtain information on the QTL of SL, SW and ST and their corresponding ratios—length-to-width (SLW), length-to-thickness (SLT) and width-to-thickness (SWT)—in soybean, including main-effect QTL (M-QTL), environment-by-QTL interaction, cytoplasm-by-QTL interaction and the gene action of identified QTL.

Materials and methods

Mapping population

The direct and reciprocal crosses were made between Lishuizhongzihuang (LSZZH) and Nannong 493-1 (N493) in 2005 because of highly significant difference between the two parents for 100-seed weight (13 g for LSZZH and 20 g for N493, $P < 0.01$). Two hundred and forty-four F_2 plants of the direct cross and 260 F_2 plants of the reciprocal cross were produced in 2006. All the F_2 plants of the above two crosses were selfed to develop $F_{2:3}$, $F_{2:4}$ and $F_{2:5}$ families, which were used in this study.

Trait evaluation

All the $F_{2:3}$, $F_{2:4}$ and $F_{2:5}$ families of the above two crosses were planted in three-row plots in a completely randomized design and evaluated at the Jiangpu experimental station at Nanjing Agricultural University in 2007, 2008 and 2009, respectively. Plots were 1.5 m wide and 2 m long. Five individuals and 20 seeds in the middle row of each plot were randomly picked out to measure seed size traits—SL, SW and ST—as well as seed shape traits—SLW, SLT and SWT ratios—by digital vernier caliper. The measurement of each seed trait was averaged over 20 seeds.

Simple and partial correlation analysis

Simple and partial correlation analysis among the above traits in soybean was applied. All statistical tests were two sided. $P < 0.05$ was considered statistically significant. The SAS 9.20 statistical package was used.

DNA extraction, PCR reaction and polymorphism detection

Approximately 0.3 g fresh leaves obtained from each F_2 plant of the above two crosses in 2006 was used to extract genomic DNA using the cetyl-trimethyl- ammonium bromide method as described by Lipp et al. (1999). To screen for polymorphisms between soybean cultivars LSZZH and N493, 972 simple sequence repeat (SSR) primer pairs were examined. Primer sequences were obtained from the soybean database Soybase (http://www.ncbi.nlm.nih.gov). PCR was performed in a volume of 15 μl containing 5 μl template DNA (20 ng/μl), 3 μl primers, 1.5 μl 10 × PCR buffer (with 15 mM/l Mg^{2+}), 0.2 μl dNTP (10 mM/l), 0.15 μl *Taq* enzyme (5 U/μl) and 5.15 μl ddH_2O. The PCR profile included one cycle of 95°C for 2 min; followed by 35 cycles of 94°C for 30 s, 47–55°C (depending on the specific primers) for 45 s and 72°C for 1 min; and a final extension at 72°C for 10 min. PCR products were separated on 8% non-denaturing polyacrylamide gels with a 29:1 ratio of acrylamide: bisacrylamide and then silver-stained as described by Santos et al. (1993).

Genetic linkage map construction and QTL mapping

JoinMap 4.0 (Van Ooijen 2006) was employed to construct a linkage map with a maximum recombinant fraction of

0.40 and a minimum LOD score of 3.0. The recombination frequency was converted to genetic map distance (centi-Morgan, cM) using the Kosambi mapping function (Kosambi 1944). Linkage groups were assigned to chromosomes based on the publicly available linkage map (Song et al. 2004) and soybean genome sequence (Schmutz et al. 2010), with those unassigned denoted as XA, where A represents serial numbers.

QTL analysis was performed using multi-QTL joint analysis under the framework of penalized maximum likelihood (Zhang and Xu 2005; Dou et al. 2010), along with composite interval mapping (CIM) by Cartographer v2.5 (Wang et al. 2007). In the joint analysis, the average phenotypic value of quantitative trait for the ith $F_{2:r}$ family of the jth family population of the kth cross ($i = 1, \ldots, n_k$; $j = 1, 2, 3$; $k = 1, 2$), y_{ijk}, may be described by the following model:

$$y_{ijk} = \mu + x_c c + E + G + GE + GC + \varepsilon_i \quad (1)$$

where μ is the population mean; c is cytoplasmic effect, and the dummy variable x_c is defined as $x_c = 1$ for a direct cross and $x_c = -1$ for a reciprocal cross; E is environmental effect; G is genetic effect; GE is environment-by-QTL interaction effect; GC is cytoplasm-by-QTL interaction effect; ε_i is a residual error with an assumed $N(0, \sigma^2)$ distribution; and

$$E = \sum_{j=1}^{R-1} x_{ij} r_j$$

$$G = \sum_{l=1}^{m} \left(x_{i(R-1+2l-1)} a_l + x_{i(R-1+2l)} d_l \right)$$

$$GE = \sum_{j=1}^{R-1} \sum_{l=1}^{m} \left[x_{ij} x_{i(R-1+2l-1)} (ae)_{jl} + x_{ij} x_{i(R-1+2l)} (de)_{jl} \right]$$

$$GC = \sum_{l=1}^{m} \left[x_c x_{i(R-1+2l-1)} (ac)_l + x_c x_{i(R-1+2l)} (dc)_l \right]$$

where R and m are the numbers of environments and QTL, respectively; a is additive effect; d is dominant effect; ae is additive-by-environment interaction effect; de is dominant-by-environment interaction effect; ac is additive-by-cytoplasm interaction effect; dc is dominant-by-cytoplasm interaction effect; and x is a dummy variable for various effects. The pseudomarker approach uses the multi-marker analysis with a slight modification by inserting virtual markers into all marker intervals >5 cM. Because of incomplete genotypic information in the real data analysis, multiple permutations for incomplete marker genotypes (Sen and Churchill 2001) were adopted to simulate the incomplete genotypes. This required multiple analyses of the data, one for each imputed dataset. Although 10–20 imputed datasets may suffice (Sen and Churchill 2001; Xu and Jia 2007), we imputed 50 samples in this study. For each sample, the complete genotypes sampled were used to construct the design matrix for QTL effects in model (1). All the effects in model (1) were simultaneously estimated by the penalized maximum likelihood method of Zhang and Xu (2005). The samples in which the LOD statistic was greater than 2.5 were counted. Furthermore, LOD score values between 2.0 and 2.5 were used to detect suggestive QTL, as suggested by Lander and Kruglyak (1995). A QTL detected in which the ratio of the number of such samples to the total number of imputed samples (50) exceeded 10% was considered a true QTL. The QTL position is an average weighted by the total genetic variance of QTL detected.

In the analysis with CIM for the jth family population of the kth cross, the standard model (Model 6), which takes forward stepwise regression with backward elimination, was adopted with a walk speed of 1 cM to search for QTL and identify cofactors. The window size was set at 10 cM, and the five background markers with the highest P value were used as cofactors to control the genetic background for each trait. A LOD score threshold of 2.5 was used to declare the presence of a putative QTL in a given genomic region. QTL confidence intervals (90–95%) were set as map intervals corresponding to one LOD decline on either side of the peak. The phenotypic variation explained (PVE) by QTL (it is similar to heritability of one QTL), and the additive and dominant effects of each QTL for all traits were calculated as well.

Results

Construction of genetic linkage map in soybean

A total of 972 SSR primer pairs covering the whole genome were used to screen for polymorphisms between the soybean cultivars LSZZH and N493. Of these, 150 primer pairs showed polymorphisms among the two parents and all the F_2 plants. Thus, 15.4% of the SSR primers revealed polymorphic bands among the parents and the F_2 plants. By using JoinMap 4.0, 113 SSR markers were mapped into 34 linkage groups, which were assigned to 19 chromosomes except for four genetic linkage maps. The total length of the linkage maps was 1,557.85 cM, with an average marker spacing of 13.79 cM.

Phenotypic evaluations

Mean value, standard deviation, range, skewness and kurtosis for each trait measured in the parent, $F_{2:3}$, $F_{2:4}$ and $F_{2:5}$ populations for 3 years of growing seasons were calculated (Table 1). There was evidence to indicate that the two average values of the parents for each trait were

Table 1 Phenotypic variation of seed traits in 504 $F_{2:3}$, $F_{2:4}$ and $F_{2:5}$ families from the soybean cross of Lishuizhongzihuang (P_1) by Nannong 493-1 (P_2)

Year	Trait	P_1	P_2	Population	Direct cross ($n = 244$)						Reciprocal cross ($n = 260$)					
					Maximum	Minimum	Mean	SD	Kurtosis	Skewness	Maximum	Minimum	Mean	SD	Kurtosis	Skewness
2007	SL	7.51 ± 0.27	8.62 ± 0.27	$F_{2:3}$	9.42	7.21	8.45	0.37	−0.01	−0.03	9.17	7.66	8.46	0.31	−0.12	−0.11
	SW	6.48 ± 0.21	7.36 ± 0.26		7.97	6.13	7.21	0.26	1.09	−0.28	7.84	6.56	7.24	0.24	−0.09	−0.22
	ST	5.31 ± 0.27	6.37 ± 0.38		6.58	5.28	5.95	0.25	0.01	−0.22	6.74	5.23	5.96	0.23	0.25	−0.09
	SLW	1.16 ± 0.04	1.17 ± 0.04		1.26	1.11	1.17	0.03	−0.24	0.21	1.23	1.10	1.17	0.03	−0.39	0.04
	SLT	1.42 ± 0.08	1.36 ± 0.09		1.61	1.31	1.42	0.06	0.21	0.35	1.62	1.27	1.42	0.06	0.24	0.28
	SWT	1.22 ± 0.04	1.16 ± 0.05		1.32	1.12	1.21	0.04	0.21	0.06	1.32	1.12	1.22	0.03	−0.19	0.01
	SL	7.32 ± 0.10	8.41 ± 0.10		8.79	6.76	7.67	0.30	0.61	0.15	8.43	6.97	7.67	0.26	−0.01	0.18
	SW	6.51 ± 0.09	7.42 ± 0.13		7.52	6.06	6.79	0.23	0.44	0.06	7.30	6.32	6.80	0.20	−0.33	0.03
	ST	5.52 ± 0.11	6.26 ± 0.11		6.41	4.97	5.64	0.22	0.65	0.08	6.27	5.14	5.66	0.20	−0.21	0.17
2008	SLW	1.12 ± 0.02	1.13 ± 0.02	$F_{2:4}$	1.20	1.06	1.13	0.03	−0.26	0.25	1.20	1.07	1.13	0.02	0.05	0.13
	SLT	1.33 ± 0.02	1.34 ± 0.03		1.52	1.23	1.36	0.05	−0.10	0.35	1.48	1.21	1.36	0.04	0.09	0.10
	SWT	1.18 ± 0.02	1.19 ± 0.02		1.29	1.12	1.20	0.03	0.02	0.12	1.29	1.12	1.20	0.03	0.44	0.28
	SL	7.25 ± 0.08	8.48 ± 0.11		8.64	7.02	7.81	0.29	−0.39	0.16	8.59	6.97	7.84	0.27	0.19	−0.21
	SW	6.32 ± 0.08	7.54 ± 0.10		7.64	5.93	6.76	0.21	1.62	0.29	7.40	6.13	6.80	0.22	0.09	−0.12
	ST	5.49 ± 0.10	6.47 ± 0.12		6.33	4.91	5.80	0.22	0.90	−0.30	6.53	5.14	5.82	0.22	0.31	−0.07
2009	SLW	1.15 ± 0.02	1.12 ± 0.02	$F_{2:5}$	1.25	1.08	1.16	0.03	−0.18	0.20	1.23	1.07	1.15	0.03	0.16	0.25
	SLT	1.32 ± 0.03	1.31 ± 0.03		1.52	1.24	1.35	0.05	−0.20	0.39	1.48	1.20	1.35	0.05	0.07	0.19
	SWT	1.15 ± 0.02	1.17 ± 0.03		1.27	1.10	1.17	0.03	0.68	0.60	1.25	1.10	1.17	0.03	−0.09	0.03

See "Introduction" for trait abbreviation definitions. The same is true for the latter tables

SD standard deviation

Table 2 Simple and partial correlation coefficients for seed traits in soybean

Trait		Direct cross						Reciprocal cross					
		SL	SW	ST	SLW	SLT	SWT	SL	SW	ST	SLW	SLT	SWT
$F_{2:3}$	SL		0.793**	0.496**	0.624**	0.546**	0.262**		0.786**	0.483**	0.452**	0.449**	0.243**
	SW	0.871**		0.717**	0.098	0.186**	0.206**	0.873**		0.707**	−0.196**	0.017	0.180**
	ST	−0.195**	0.651**		−0.075	−0.441**	−0.534**	−0.155*	0.616**		−0.252**	−0.564**	−0.567**
	SLW	0.590**	−0.436**	−0.043		0.728**	0.221**	0.574**	−0.379**	−0.161*		0.686**	0.124
	SLT	0.123	−0.205**	0.220**	0.728**		0.829**	0.185**	−0.308**	0.330**	0.699**		0.807**
	SWT	−0.183**	0.452**	−0.621**	−0.606**	0.901**		−0.208**	0.499**	−0.679**	−0.612**	0.917**	
$F_{2:4}$	SL		0.723**	0.509**	0.534**	0.481**	0.198**		0.777**	0.554**	0.534**	0.420**	0.105
	SW	0.828**		0.805**	−0.092	−0.003	0.106	0.902**		0.774**	−0.118	−0.049	0.040
	ST	−0.214**	0.724**		−0.215**	−0.468**	−0.504**	−0.140*	0.553**		−0.167**	−0.521**	−0.600**
	SLW	0.347**	−0.073	−0.299**		0.791**	0.232**	0.592**	−0.506**	0.018		0.730**	0.114
	SLT	0.227**	−0.413**	0.439**	0.834**		0.778**	0.357**	−0.354**	0.125	0.541**		0.762**
	SWT	−0.257**	0.591**	−0.713**	−0.761**	0.943**		−0.362**	0.548**	−0.559**	−0.443**	0.892**	
$F_{2:5}$	SL		0.664**	0.366**	0.595**	0.555**	0.278**		0.739**	0.506**	0.413**	0.419**	0.217**
	SW	0.112		0.746**	−0.122	0.019	0.149*	0.750**		0.802**	−0.308**	−0.142*	0.104
	ST	0.065	0.983**		−0.275**	−0.528**	−0.516**	−0.240**	0.822**		−0.366**	−0.570**	−0.510**
	SLW	0.117	−0.971**	0.955**		0.796**	0.222**	0.393**	−0.032	−0.293**		0.781**	0.162*
	SLT	0.060	0.959**	−0.976**	0.982**		0.740**	0.149^{+}	−0.392**	0.450**	0.850**		0.743**
	SWT	−0.008	−0.075	0.083	−0.287**	0.289**		-0.214**	0.657**	-0.782**	-0.731**	0.908**	

*, ** Significant at 0.05 and 0.01 levels, respectively. The simple and partial correlation coefficients are listed in the top right and bottom left corners, respectively

significantly different ($P < 0.05$). All traits from six datasets exhibited continuous distribution in the above populations and almost showed a normal distribution with skewness and kurtosis statistics, typical of quantitative traits.

Simple and partial correlations among seed size traits based upon $F_{2:3}$, $F_{2:4}$ and $F_{2:5}$ family means from 2007 to 2009 are given in the top right and bottom left corners in Table 2, respectively. In a simple correlation analysis, there was a significant positive correlation between any two traits among seed size traits (SL, SW and ST) but a significant negative correlation between thickness and each seed shape trait (SLW, SLT and SWT). In the partial correlation analysis, significant positive correlations for most trait pairs, except for SL and ST (or SWT), SW and SLW (or SLT) and SWT and SL (or ST, or SLW), were observed.

Mapping M-QTL

A total of 121 M-QTL: 21 for SL, 21 for SW, 22 for ST, 24 for SLW, 18 for SLT and 15 for SWT were detected by the multi-QTL joint analysis of the six datasets (Supplement Table A1), whereas a total of 67 M-QTL: 10 for SL, 12 for SW, 12 for ST, 19 for SLW, 8 for SLT and 7 for SWT were identified by CIM in the direct and reciprocal cross populations of the $F_{2:3}$, $F_{2:4}$ and $F_{2:5}$ families. A summary of all 52 common M-QTL detected, including marker interval, position and LOD score, additive and dominant effects and percentage of the PVE by the QTL is shown in Table 3 and Fig. 1.

Eight common SL M-QTL, with heritabilities of 4.51–14.94%, were simultaneously identified by both the joint analysis and CIM and mapped to chromosomes 6, 10, 16, 17, 18 and 19. Of these QTL, seven M-QTL had negative additive effects, −0.081 to −0.037, which indicates that N493 contributed one allele leading to an increase in SL. Two M-QTL, *qSL-10-2* and *qSL-18*, were identified in more than one environment by CIM, and the heritabilities of *qSL-10-2* and *qSL-10-3* were greater than 10%. In addition, the heritabilities of *qSL-2-2* and *qSL-14-1*, which were detected by CIM but not by the joint analysis, were 16.10 and 24.87%, respectively.

Nine common SW M-QTL, located on chromosomes 2, 3, 5, 6, 7, 10, 14, 18 and 20, were identified simultaneously by both the joint analysis and CIM, with heritabilities ranging from 5.61 to 28.21%. Of these QTL, seven M-QTL had negative additive effects, −0.054 to −0.011, which indicates that N493 contributed one allele leading to an increase in SW. Two QTL, *qSW-6-1* and *qSW-20*, were detected in more than one environment by CIM, and *qSW-6-1* was mapped simultaneously in the $F_{2:3}$, $F_{2:4}$ and $F_{2:5}$ family populations. The heritabilities of *qSW-2-1*, *qSW-6-1*, *qSW-14-2* and *qSW-20* were greater than 10%.

Eight common ST M-QTL, located on chromosomes 3, 5, 6, 7, 18 and 20, were found simultaneously by the joint

Table 3 Common main-effect QTL identified by multi-QTL joint analysis and composite interval mapping

Trait	QTL[a]	Chr	Marker interval	Multi-QTL joint analysis					Composite interval mapping					
				Position (cM)	LOD ± SD	Additive ± SD	Dominant ± SD	Ratio (%)[b]	Position (cM)	LOD	Additive	Dominant	PVE (%)[c]	Population source
SL	*qSL-6*	C2-1, 6	satt640-satt422	14.3	5.54 ± 1.36	−0.048 ± 0.006	−0.009 ± 0.012	90	8.0	2.64	−0.055	−0.079	7.07	Direct cross (DC)($F_{2:5}$)
	qSL-10-1	O-1, 10	satt173-satt094	36.7	3.56 ± 0.69	0.028 ± 0.011	−0.038 ± 0.011	16	41.4	2.92	0.078	−0.141	5.38	DC($F_{2:4}$)
	qSL-10-2	O-2, 10	satt331-satt592	0.7	12.69 ± 6.85	−0.081 ± 0.016	−0.016 ± 0.009	92	0–5.0	2.71–5.89	−0.133 to −0.104	−0.008 to 0.012	4.51–10.60	DC($F_{2:4}$), Reciprocal cross (RC)($F_{2:3}$,$F_{2:4}$)
	qSL-10-3	O-2,10	satt592-sat_274	31.5	3.48 ± 0.98	−0.045 ± 0.007	0.008 ± 0.015	40	26.1–49.9	3.77–3.83	−0.092 to 0.071	−0.099 to 0.084	6.63–14.94	DC($F_{2:5}$), RC($F_{2:5}$)
	qSL-16	J,16	sat_224-sat_394	85.4	3.04 ± 0.40	−0.037 ± 0.005	0.007 ± 0.015	10	74.5–76.7	2.55	−0.217	0.134	5.79	DC($F_{2:3}$)
	qSL-17-1	D2-1, 17	satt226-sat_354	76.5	3.82 ± 1.40	−0.043 ± 0.009	0.000 ± 0.009	20	75.4	3.09	−0.125	0.018	9.20	DC($F_{2:5}$)
	qSL-18	G, 18	satt309-satt688	13.1	3.96 ± 0.88	−0.046 ± 0.009	0.019 ± 0.024	80	0–12.1	2.81–3.19	−0.095 to 0.075	−0.133 to −0.013	4.94–5.16	DC($F_{2:5}$), RC($F_{2:3}$)
	qSL-19	L-2, 19	satt166-satt527	4.8	3.56 ± 0.92	−0.038 ± 0.005	0.014 ± 0.011	48	0	2.92	−0.105	0.140	6.64	DC($F_{2:4}$)
SW	*qSW-2-1*	D1b-1, 2	satt157-sat_254	35.5	4.21 ± 1.20	−0.028 ± 0.006	0.020 ± 0.010	86	19.0	3.71	−0.125	0.144	17.85	DC($F_{2:4}$)
	qSW-3-2	N-2, 3	satt255-satt237	9.5	3.50 ± 0.56	0.031 ± 0.002	0.003 ± 0.009	24	0	3.42	0.082	−0.002	5.66	RC($F_{2:5}$)
	qSW-5-2	A1, 5	satt449-sat_356	66.7	5.65 ± 1.39	−0.038 ± 0.005	0.005 ± 0.009	78	61.9	3.44	−0.096	0.026	9.80	DC($F_{2:4}$)
	qSW-6-1	C2-1, 6	satt640-satt422	23.1	10.53 ± 5.83	−0.054 ± 0.011	−0.001 ± 0.009	88	9–25.0	2.59–6.42	−0.103 to −0.043	−0.057 to 0.074	5.70–11.38	DC(F2:5), RC($F_{2:3}$,$F_{2:4}$,$F_{2:5}$)
	qSW-7-2	M, 7	satt323-satt220	54.0	5.31 ± 1.38	−0.039 ± 0.005	0.010 ± 0.006	62	53.7	3.59	−0.097	0.027	6.53	DC($F_{2:4}$)
	qSW-10-2	O-2, 10	satt592-sat_274	21.4	5.14 ± 1.19	−0.036 ± 0.004	−0.006 ± 0.007	16	18.0	3.12	−0.025	−0.089	7.66	RC($F_{2:5}$)
	qSW-14-2	B2-2, 14	satt355-satt070	7.5	3.23 ± 0.67	0.022 ± 0.009	−0.002 ± 0.025	10	4.0	2.63	0.023	0.165	28.21	RC($F_{2:5}$)
	qSW-18	G, 18	satt309-satt688	1.0	2.92 ± 0.33	−0.011 ± 0.004	−0.035 ± 0.003	20	1.0	3.09	0.029	−0.100	5.61	DC($F_{2:5}$)
	qSW-20	I-2, 20	sat_418-sat_419	35.1	6.21 ± 2.70	−0.041 ± 0.007	0.001 ± 0.008	84	24.0–28.0	3.24–4.09	−0.117 to −0.076	−0.031 to 0.016	9.30–11.43	RC($F_{2:4}$,$F_{2:5}$)
ST	*qST-3-2*	N-2, 3	satt255-satt237	14.2	4.91 ± 0.70	0.038 ± 0.003	−0.004 ± 0.007	12	4.0	3.34	0.096	−0.015	6.35	RC($F_{2:5}$)
	qST-3-4	N-2, 3	satt339-satt022	30.8	5.60 ± 1.14	0.040 ± 0.004	−0.003 ± 0.013	10	55.2	2.68	0.085	−0.058	5.66	RC($F_{2:4}$)
	qST-5-2	A1, 5	satt449-sat_356	74.2	8.19 ± 4.00	−0.047 ± 0.009	0.016 ± 0.007	84	44.1–63.9	2.86–5.82	−0.197 to −0.073	−0.058 to 0.062	9.79–23.69	DC($F_{2:3}$,$F_{2:4}$), RC($F_{2:3}$)
	qST-6-1	C2-1, 6	satt640-satt422	4.9	3.66 ± 0.82	−0.028 ± 0.003	−0.010 ± 0.005	76	15.0	3.72	−0.084	−0.014	7.84	RC($F_{2:5}$)
	qST-7-1[d]	M, 7	satt150-sat_258	8.1	3.18 ± 0.43	−0.023 ± 0.010	0.015 ± 0.020	10	7.0	2.91	−0.105	0.044	8.98	DC($F_{2:5}$)
	qST-7-2	M, 7	sat_258-satt463	30.8	4.28 ± 0.91	−0.032 ± 0.004	0.019 ± 0.010	30	25.6	4.34	−0.113	0.142	11.29	DC($F_{2:4}$)
	qST-18	G, 18	satt309-satt688	3.6	3.28 ± 0.67	−0.002 ± 0.004	−0.041 ± 0.004	16	0	3.01	0.017	−0.094	5.43	DC($F_{2:5}$)
	qST-20-2	I-2, 20	sat_418-sat_419	22.5	3.74 ± 0.79	−0.030 ± 0.004	−0.004 ± 0.011	60	5.0	2.93	−0.081	−0.006	6.01	RC($F_{2:5}$)
SLW	*qSLW-1-1*	D1a-2, 1	AW781285-satt077	58.3	5.39 ± 1.33	−0.040 ± 0.008	0.024 ± 0.009	88	54.1–58.9	2.73–4.32	−0.010 to −0.007	0.004–0.006	2.93–3.93	Joint direct and reciprocal crosses (JDRC)($F_{2:4}$,$F_{2:5}$)
	qSLW-2	D1b-2, 2	satt266-satt282	18.2	4.85 ± 1.88	−0.036 ± 0.011	0.026 ± 0.015	44	22.0–33.0	2.51–3.45	−0.010 to −0.008	0.004–0.013	4.38–7.41	(JDRC)($F_{2:4}$), RC($F_{2:4}$,$F_{2:5}$)
	qSLW-5-1	A1, 5	sat_344-satt449	11.5	3.79 ± 1.03	0.036 ± 0.005	−0.004 ± 0.011	22	15.0–20.0	2.57–2.83	0.004–0.012	−0.003 to 0.017	7.78–14.56	(JDRC)($F_{2:4}$,$F_{2:5}$), C($F_{2:4}$,$F_{2:5}$)
	qSLW-5-2	A1, 5	satt449-sat_356	71.4	5.53 ± 2.09	0.044 ± 0.008	−0.001 ± 0.010	88	54.9–62.9	2.59–3.07	0.005–0.006	0.004–0.008	4.78–5.43	(JDRC)($F_{2:3}$,$F_{2:4}$)
	qSLW-6	C2-1, 6	satt640-satt422	22.3	4.19 ± 1.15	0.036 ± 0.005	−0.013 ± 0.008	90	11.0–25.0	2.61–5.82	0.008–0.014	−0.002 to 0	2.83–9.59	(JDRC)($F_{2:5}$), RC($F_{2:4}$,$F_{2:5}$)
	qSLW-10-1	O-2, 10	satt331-satt592	1.2	5.90 ± 1.80	−0.052 ± 0.007	−0.008 ± 0.012	82	1.0	2.83–2.87	−0.004 to −0.002	−0.009 to −0.003	2.41–5.18	(JDRC)($F_{2:4}$), DC($F_{2:4}$)
	qSLW-10-2	O-2, 10	satt592-sat_274	48.7	5.52 ± 1.86	−0.046 ± 0.008	0.016 ± 0.012	82	9.0–49.0	3.88–9.59	−0.016 to −0.010	0.003–0.023	6.03–30.30	(JDRC)($F_{2:3}$, $F_{2:4}$, $F_{2:5}$), DC($F_{2:3}$, $F_{2:4}$, $F_{2:5}$), RC($F_{2:3}$,$F_{2:4}$)
	qSLW-13-1	F, 13	satt649-satt269	0.4	6.97 ± 1.27	0.050 ± 0.005	0.007 ± 0.008	96	0–1.0	2.53–4.13	0.006–0.007	−0.002 to 0.003	2.21–3.98	(JDRC)($F_{2:3}$, $F_{2:4}$,$F_{2:5}$), RC($F_{2:4}$)

Table 3 continued

Trait	QTL[a]	Chr	Marker interval	Multi-QTL joint analysis					Composite interval mapping					
				Position (cM)	LOD ± SD	Additive ± SD	Dominant ± SD	Ratio (%)[b]	Position (cM)	LOD	Additive	Dominant	PVE (%)[c]	Population source
SLW	*qSLW-16-1*	J, 16	sat_224-sat_394	107.2	4.09 ± 0.91	0.005 ± 0.007	−0.052 ± 0.006	50	100.7	2.86	0.005	−0.019	10.26	(JDRC)($F_{2:5}$)
	qSLW-17-2	D2-1, 17	sat_292-satt615	25.9	3.92 ± 0.79	−0.038 ± 0.004	0.004 ± 0.005	10	29.7	2.52	−0.007	−0.006	5.76	DC($F_{2:5}$)
	qSLW-17-3	D2-1, 17	satt226-sat_354	73.4	4.66 ± 1.36	−0.040 ± 0.007	−0.003 ± 0.013	46	75.4	2.65	−0.009	0.004	4.14	(JDRC)($F_{2:5}$)
	qSLW-17-4	D2-2, 17	satt672-satt413	5.6	8.21 ± 3.59	−0.054 ± 0.009	0.015 ± 0.011	44	4.0	2.76–2.95	−0.013 to −0.019	0.003–0.007	5.19–7.91	DC($F_{2:3}$), RC($F_{2:4}$)
	qSLW-17-5	D2-2, 17	satt413-satt256	19.4	7.73 ± 2.69	−0.047 ± 0.013	0.017 ± 0.019	66	16.0–24.0	2.66–6.13	−0.013 to −0.009	0.003–0.012	4.93–9.87	(JDRC)($F_{2:3}$, $F_{2:4}$,$F_{2:5}$), DC($F_{2:4}$), RC($F_{2:3}$,$F_{2:4}$,$F_{2:5}$)
	qSLW-18	G, 18	satt309-satt688	12.8	4.62 ± 1.09	−0.038 ± 0.005	0.011 ± 0.011	98	11.0–12.0	2.80–4.04	−0.010 to −0.006	−0.004 to −0.001	4.73–6.71	RC($F_{2:3}$,$F_{2:4}$)
	qSLW-19	L-1, 19	satt652-satt284	5.5	3.25 ± 0.71	−0.032 ± 0.003	−0.003 ± 0.012	20	0.0	2.72	−0.008	0.002	4.43	DC($F_{2:4}$)
	qSLW-X3	X3	satt260-satt382	0.9	5.39 ± 2.13	−0.035 ± 0.011	0.029 ± 0.012	60	0.0	2.51–4.13	−0.010 to −0.007	0.002–0.014	4.62–9.02	(JDRC)($F_{2:4}$), DC($F_{2:4}$,), RC($F_{2:4}$)
SLT	*qSLT-10-1*	O-2, 10	satt331-satt592	3.3	5.04 ± 2.12	−0.088 ± 0.014	−0.006 ± 0.027	46	10.0	2.76–3.07	−0.019 to −0.016	0.007–0.014	4.25–5.12	RC($F_{2:3}$, $F_{2:4}$)
	qSLT-10-2	O-2, 10	satt592-sat_274	17.7	6.17 ± 3.61	−0.086 ± 0.026	0.051 ± 0.030	42	0–49.0	2.53–2.57	−0.020 to −0.018	0.013–0.019	4.51–4.69	DC($F_{2:4}$, $F_{2:5}$)
	qSLT-13	F, 13	satt649-satt269	0.0	9.85 ± 1.40	0.121 ± 0.009	0.003 ± 0.013	96	0–1.0	2.82–3.16	0.014–0.025	−0.013 to 0	4.94–6.64	DC($F_{2:3}$, $F_{2:4}$)
	qSLT-20	I-1, 20	satt419-satt354	1.1	9.79 ± 3.24	0.119 ± 0.018	0.014 ± 0.021	86	0	3.39–4.91	0.014–0.024	−0.006 to 0.018	7.57–12.78	DC($F_{2:3}$, $F_{2:4}$), RC($F_{2:3}$)
	qSLT-X3	X3	satt260-satt382	3.1	4.45 ± 1.07	−0.054 ± 0.017	0.074 ± 0.028	62	0	2.81–3.13	−0.020 to −0.017	0.016–0.026	5.49–8.30	DC($F_{2:4}$,), RC($F_{2:3}$)
SWT	*qSWT-3-1*	N-1, 3	satt631-satt683	8.1	9.02 ± 4.95	−0.065 ± 0.016	0.020 ± 0.013	42	8.0	3.13	−0.011	0.001	7.15	DC($F_{2:4}$)
	qSWT-3-2	N-1, 3	satt683-sat_280	19.6	8.78 ± 3.95	−0.065 ± 0.013	−0.002 ± 0.013	48	47.0	3.57	−0.020	0.035	48.92	RC($F_{2:5}$)
	qSWT-6	C2-1, 6	satt640-satt422	25.2	8.20 ± 1.32	−0.061 ± 0.005	0.023 ± 0.007	98	25.0	3.97	−0.015	0.004	6.71	RC($F_{2:3}$)
	qSWT-13-1	F, 13	satt649-satt269	0.7	5.90 ± 1.44	0.055 ± 0.007	0.003 ± 0.006	100	0.0	2.73	0.008	−0.001	4.85	DC($F_{2:4}$)
	qSWT-15	E, 15	satt045-satt263	7.1	3.15 ± 0.47	0.035 ± 0.004	−0.021 ± 0.011	20	11.1	3.60	0.015	−0.008	5.55	RC($F_{2:3}$)
	qSWT-20-1	I-1, 20	satt419-satt354	1.3	7.27 ± 2.49	0.063 ± 0.009	0.007 ± 0.010	92	0–2.0	2.58–8.04	0.005–0.015	0.001–0.008	7.08–10.70	DC($F_{2:4}$), RC($F_{2:3}$,$F_{2:5}$)

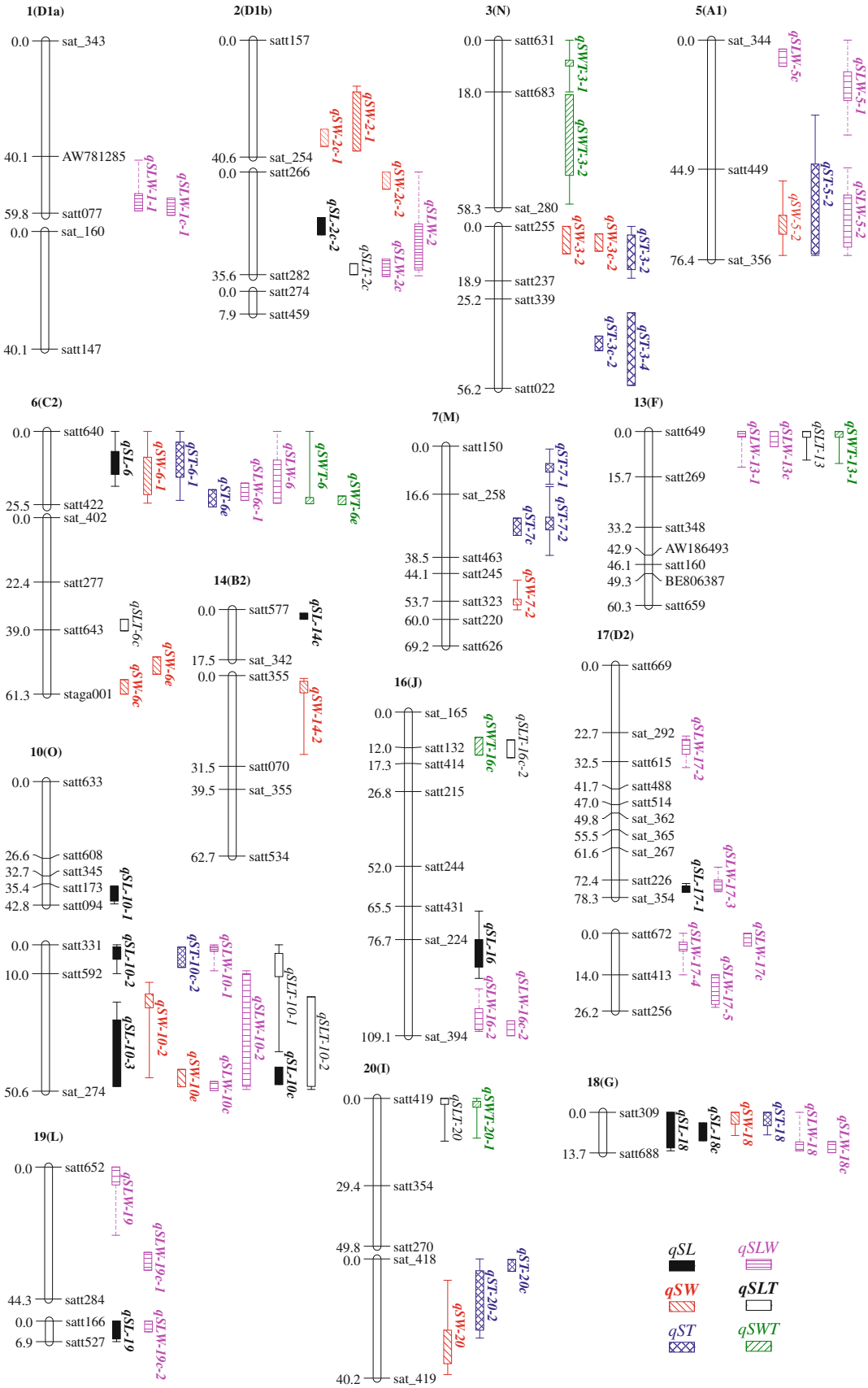
LOD log of odd, *SD* standard deviation

[a] QTL nomenclature is adopted according to the method used in rice (McCouch et al. 1997), starting with 'q', followed by an abbreviation of the trait name (for example *SL* seed length, *SW* seed width, and *ST* seed thickness) and the name of the chromosome, followed by the number of QTL affecting the trait on the chromosome

[b] The ratio is the frequency of such samples which had the LOD statistic greater than 2.5 to the total number of imputed samples (50)

[c] PVE (%): phenotypic variation explained by QTL

[d] Suggestive QTL. The same is true for the latter tables


1(D1a)
0.0 sat_343
40.1 AW781285
59.8 satt077
0.0 sat_160
40.1 satt147
2(D1b)
0.0 satt157
40.6 sat_254
0.0 satt266
35.6 satt282
0.0 satt274
7.9 satt459
3(N)
0.0 satt631
18.0 satt683
58.3 sat_280
0.0 satt255
18.9 satt237
25.2 satt339
56.2 satt022
5(A1)
0.0 sat_344
44.9 satt449
76.4 sat_356
6(C2)
0.0 satt640
25.5 satt422
0.0 sat_402
22.4 satt277
39.0 satt643
61.3 staga001
7(M)
0.0 satt150
16.6 sat_258
38.5 satt463
44.1 satt245
53.7 satt323
60.0 satt220
69.2 satt626
13(F)
0.0 satt649
15.7 satt269
33.2 satt348
42.9 AW186493
46.1 satt160
49.3 BE806387
60.3 satt659
14(B2)
0.0 satt577
17.5 sat_342
0.0 satt355
31.5 satt070
39.5 sat_355
62.7 satt534
16(J)
0.0 sat_165
12.0 satt132
17.3 satt414
26.8 satt215
52.0 satt244
65.5 satt431
76.7 sat_224
109.1 sat_394
17(D2)
0.0 satt669
22.7 sat_292
32.5 satt615
41.7 satt488
47.0 satt514
49.8 sat_362
55.5 sat_365
61.6 sat_267
72.4 satt226
78.3 sat_354
0.0 satt672
14.0 satt413
26.2 satt256
10(O)
0.0 satt633
26.6 satt608
32.7 satt345
35.4 satt173
42.8 satt094
0.0 satt331
10.0 satt592
50.6 sat_274
20(I)
0.0 satt419
29.4 satt354
49.8 satt270
0.0 sat_418
40.2 sat_419
18(G)
0.0 satt309
13.7 satt688
19(L)
0.0 satt652
44.3 satt284
0.0 satt166
6.9 satt527
qSL
qSLW
qSW
qSLT
qST
qSWT

◀ **Fig. 1** Chromosome locations of QTL associated with seed length (*SL*), width (*SW*), thickness (*ST*) and their corresponding ratios (length-to-width, *SLW*; length-to-thickness, *SLT*; and width-to-thickness, *SWT*) in the $F_{2:3}$, $F_{2:4}$ and $F_{2:5}$ populations from the direct and reciprocal crosses of Lishuizhongzihuang × Nannong 493-1 soybeans (*Glycine max* L. Merr.), with the software of Mapchart 2.2 (Voorrips 2002). Positions of loci are given in centi-Morgan. Fifty-two main-effect QTL, 4 environmental interaction QTL and 28 nucleo-cytoplasmic interaction QTL are shown for SL, SW, ST, SLW, SLT and SWT

analysis and CIM, with heritabilities ranging from 5.43 to 23.69%. Of these QTL, six had negative additive effects, −0.047 to −0.002, which means that N493 contributed one allele leading to an increase in ST. One QTL, *qST-5-2*, was confirmed in more than one environment by CIM. The heritabilities for *qST-5-2* and *qST-7-2* were greater than 10%.

Sixteen common SLW M-QTL, located on chromosomes 1, 2, 5, 6, 10, 13, 16, 17, 18, 19 and X3, were identified simultaneously by the joint analysis and CIM, with heritabilities ranging from 2.21 to 30.30%. Of these QTL, 11 had negative additive effects, −0.054 to −0.032, which means that N493 contributed one allele leading to an increase in seed thickness. Ten QTL, *qSLW-1-1*, *qSLW-2*, *qSLW-5-1*, *qSLW-5-2*, *qSLW-6*, *qSLW-10-2*, *qSLW-13-1*, *qSLW-17-4*, *qSLW-17-5* and *qSLW-18*, were found in more than one environment by CIM, and three of them, *qSLW-10-2*, *qSLW-13-1* and *qSLW-17-5*, were mapped simultaneously in the $F_{2:3}$, $F_{2:4}$ and $F_{2:5}$ family populations. The heritabilities for the three QTL, *qSLW-5-1*, *qSLW-10-2* and *qSLW-16-1*, were greater than 10%.

Five common SLT M-QTL, located on chromosomes 10, 13, 20 and X3, with heritabilities ranging from 4.25 to 12.78%, were detected simultaneously by the joint analysis and CIM. Of these QTL, three had a negative additive effect, −0.088 to −0.054, which means that N493 contributed one allele leading to an increase in seed thickness. Five QTL, *qSLT-10-1*, *qSLT-10-2*, *qSLT-13*, *qSLT-20* and *qSLT-X3*, were mapped in more than one environment by CIM. One QTL, *qSLT-20*, had a heritability greater than 10%.

Six common SWT M-QTL, located on chromosomes 3, 6, 13, 15 and 20, with heritabilities ranging from 4.85 to 48.92%, were identified simultaneously by the joint analysis and CIM. Of these QTL, three had negative additive effects, −0.065 to −0.061, which means that N493 contributed one allele leading to an increase in seed thickness. One QTL, *qSWT-20-1*, was mapped simultaneously in all populations, and *qSWT-3-2* and *qSWT-20-1* had a high heritability, greater than 10%, particularly *qSWT-3-2* (48.92%).

In conclusion, the following are five common M-QTL in all three populations: *qSW-6-1, qSLW-10-2, qSLW-13-1, qSLW-17-5* and *qSWT-20-1*. The following five M-QTL had a heritability of greater than 20%: *qSL-14-1, qSW-14-2, qST-5-2, qSLW-10-2* and *qSWT-3-2*.

Cytoplasmic effect and cytoplasm-by-QTL interaction

Significant cytoplasmic effects, estimated to be between −0.151 and −0.063, were evident for all traits except for SLW. A total of 92 cytoplasm-by-QTL interactions: 10, 11, 11, 23, 18 and 19 interactions for SL, SW, ST, SLW, SLT and SWT, respectively, were detected (Supplement Table A2). Among these interactions, 28 were consistent with M-QTL detected by CIM (Table 4). Among the 28 interactions, three QTL, *qSL-10-3, qSLW-10-2* and *qSLW-X3*, were detected in the direct and reciprocal crosses but had different effects and heritabilities across the direct and reciprocal crosses, and the others were detected in only the direct or the reciprocal cross.

Environmental effect and environment-by-QTL interaction

Significant environmental effects were observed for all the traits (Supplement Table A4). A total of 8 environment-by-QTL interactions: 2, 2, 2 and 2 interactions for the SL, SW, ST and SWT, respectively, were identified (Supplement Table A3). Among these interactions, four were consistent with M-QTL identified by CIM (Table 5) and mapped only in one environment.

Discussion

In mapping QTL for quantitative traits, CIM is a widely used approach because it solves the problem of multiple QTL residing in the same linkage group. However, it can only detect QTL in a single population. To compensate for this shortcoming, MJA is often used. There are some advantages in using MJA on the $F_{2:3}$, $F_{2:4}$ and $F_{2:5}$ populations over CIM in one population. First, environment-related effects can be identified, e.g., the six environmental effects and eight environment-by-QTL interactions that were identified in this study. Second, cytoplasm-related effects can be found, e.g., the five cytoplasmic effects and 92 cytoplasm-by-QTL interactions that were detected in this study. Third, it can validate results derived from CIM, e.g., the 52 M-QTL detected in the $F_{2:3}$, $F_{2:4}$ and $F_{2:5}$ that were further confirmed by MJA. Finally, it can detect new QTL not identified by CIM, e.g., the 69 new QTL that were found in MJA. Therefore, a multi-environment, multi-marker, multi-cross (direct and reciprocal crosses) and multi-population joint analysis was adopted to dissect the genetic architecture for seed size and shape traits in this study.

Table 4 Cytoplasm-by-QTL interactions that are consistent with main-effect QTL detected by composite interval mapping

Trait	Chr	Marker interval	Cytoplasm-by-QTL interaction						Composite interval mapping						
			QTL[a]	Position (cM)	LOD ± SD	Additive ± SD	Dominant ± SD	Ratio (%)	QTL	Position (cM)	LOD	Additive	Dominant	PVE(%)	Population source
SL	D1b-2, 2	satt266-satt282	*qSL-2c-2*	19.8	3.66 ± 1.12	0.023 ± 0.015	−0.044 ± 0.016	52	*qSL-2-2*	13.0	6.34	−0.171	0.101	16.10	Reciprocal cross (RC)($F_{2:4}$)
	O-2, 10	satt592-sat_274	*qSL-10c*	44.5	4.08 ± 0.99	−0.043 ± 0.005	−0.010 ± 0.013	24	*qSL-10-3*	26.1–49.9	3.77–3.83	−0.092 to 0.071	−0.099 to 0.084	6.63–14.94	Direct cross (DC)($F_{2:3}$,$F_{2:5}$) RC($F_{2:5}$)
	B2-1, 14	satt577-sat_342	*qSL-14c*	2.3	3.40 ± 0.70	−0.009 ± 0.013	−0.052 ± 0.003	10	*qSL-14-1*	17.0	3.37	0.046	0.048	24.87	RC($F_{2:5}$)
	G, 18	satt309-satt688	*qSL-18c*	5.9	3.76 ± 0.84	0.036 ± 0.005	−0.026 ± 0.013	42	*qSL-18*	0–12.1	2.81–3.19	−0.095 to 0.075	−0.133 to −0.013	4.94–5.16	DC($F_{2:5}$), RC($F_{2:3}$)
SW	D1b-1, 2	satt157-sat_254	*qSW-2c-1*	33.9	3.51 ± 0.83	0.002 ± 0.006	0.043 ± 0.004	52	*qSW-2-1*	19.0	3.71	−0.125	0.144	17.85	DC($F_{2:4}$)
	D1b-2, 2	satt266-satt282	*qSW-2c-2*	3.5	3.77 ± 0.96	0.026 ± 0.011	−0.023 ± 0.013	34	*qSW-2-2*	0	2.59	−0.072	0.044	5.13	RC($F_{2:4}$)
	N-2, 3	satt255-satt237	*qSW-3c-2*	6.2	3.31 ± 0.61	−0.029 ± 0.003	−0.006 ± 0.009	14	*qSW-3-2*	0	3.42	0.082	−0.002	5.66	RC($F_{2:5}$)
	C2-2, 6	satt643-staga001	*qSW-6c*	59.4	3.31 ± 0.45	−0.014 ± 0.004	−0.036 ± 0.004	28	*qSW-6-2*	59.0	3.41	−0.031	−0.095	8.63	DC($F_{2:4}$)
ST	N-2, 3	satt339-satt022	*qST-3c-2*	41.0	3.49 ± 1.01	−0.012 ± 0.011	0.037 ± 0.009	12	*qST-3-4*	55.2	2.68	0.085	−0.058	5.66	RC($F_{2:4}$)
	O-2, 10	satt331-satt592	*qST-10c-2*	3.3	3.62 ± 1.16	0.003 ± 0.001	0.044 ± 0.006	12	*qST-10-1*	2.0	3.29	−0.064	0.117	7.35	DC($F_{2:5}$)
	M, 7	sat_258-satt463	*qST-7c*	27.4	3.17 ± 0.68	−0.029 ± 0.003	0.002 ± 0.013	16	*qST-7-2*	25.6	4.34	−0.113	0.142	11.29	DC($F_{2:4}$)
	I-2, 20	sat_418-sat_419	*qST-20c*	3.9	3.54 ± 0.74	0.010 ± 0.008	0.039 ± 0.008	54	*qST-20-2*	5.0	2.93	−0.081	−0.006	6.01	RC($F_{2:5}$)
SLW	D1a-2, 1	AW781285-satt077	*qSLW-1c-1*	57.5	3.47 ± 1.17	−0.035 ± 0.006	−0.001 ± 0.009	46	*qSLW-1-1*	54.1–58.9	2.73–4.32	−0.010 to −0.007	0.004–0.006	2.93–3.93	Joint direct and reciprocal crosses (JDRC) ($F_{2:4}$,$F_{2:5}$)
	D1b-2, 2	satt266-satt282	*qSLW-2c*	33.2	3.31 ± 0.65	0.009 ± 0.005	−0.047 ± 0.004	34	*qSLW-2-1*	22.0–33.0	2.51–3.45	−0.010 to −0.008	0.004–0.013	4.38–7.41	JDRC($F_{2:4}$), RC($F_{2:4}$,$F_{2:5}$)
	A1, 5	sat_344-satt449	*qSLW-5c*	6.1	3.48 ± 1.32	0.034 ± 0.006	−0.008 ± 0.007	18	*qSLW-5-1*	15.0–20.0	2.57–2.83	0.004–0.012	−0.003 to 0.017	7.78–14.56	JDRC($F_{2:4}$,$F_{2:5}$), DC($F_{2:4}$,$F_{2:5}$)
	C2-1, 6	satt640-satt422	*qSLW-6c-1*	21.0	4.45 ± 1.15	−0.036 ± 0.006	−0.016 ± 0.011	84	*qSLW-6-1*	11.0–25.0	2.61–5.82	0.008–0.014	−0.002 to 0.000	2.83–9.59	JDRC($F_{2:5}$), RC($F_{2:4}$,$F_{2:5}$)
SLW	O-2, 10	satt592-sat_274	*qSLW-10c*	50.3	4.89 ± 1.44	−0.039 ± 0.008	0.010 ± 0.010	62	*qSLW-10-2*	9.0–49.0	3.88–9.59	−0.016 to −0.010	0.003–0.023	6.03–30.30	JDRC($F_{2:3}$,$F_{2:4}$,$F_{2:5}$), DC($F_{2:3}$, $F_{2:4}$,$F_{2:5}$), RC($F_{2:3}$, $F_{2:4}$)
	F, 13	satt649-satt269	*qSLW-13c*	2.3	3.47 ± 0.38	0.011 ± 0.004	−0.046 ± 0.003	14	*qSLW-13-1*	0–1.0	2.53–4.13	0.006–0.007	−0.002 to 0.003	2.21–3.98	JDRC($F_{2:3}$,$F_{2:4}$,$F_{2:5}$), RC($F_{2:4}$)
	J, 16	sat_224-sat_394	*qSLW-16c-2*	107.1	4.15 ± 1.35	0.013 ± 0.007	−0.051 ± 0.010	34	*qSLW-16-1*	100.7	2.86	0.005	−0.019	10.26	JDRC($F_{2:5}$)
	D2-2, 17	satt672-satt413	*qSLW-17c*	1.3	3.52 ± 0.80	−0.032 ± 0.003	0.014 ± 0.007	18	*qSLW-17-5*	4.0	2.76–2.95	−0.013 to −0.019	0.003–0.007	5.19–7.91	DC($F_{2:3}$), RC($F_{2:4}$)
	G, 18	satt309-satt688	*qSLW-18c*	12.9	6.52 ± 1.26	0.045 ± 0.005	0.016 ± 0.007	98	*qSLW-18-1*	11.0–12.0	2.80–4.04	−0.010 to −0.006	−0.004 to −0.001	4.73–6.71	RC($F_{2:3}$,$F_{2:4}$)
	L-1, 19	satt652-satt284	*qSLW-19c-1*	31.6	3.81 ± 1.22	−0.032 ± 0.009	−0.015 ± 0.014	76	*qSLW-19-1*	0.0	2.72	−0.008	0.002	4.43	DC($F_{2:4}$)
	L-2, 19	satt166-satt527	*qSLW-19c-2*	0.6	3.56 ± 0.95	0.014 ± 0.009	0.041 ± 0.008	42	*qSLW-19-2*	0.0	3.24	−0.004	0.016	7.34	DC($F_{2:5}$)
	X3	satt260-satt382	*qSLW-X3c*	5.3	3.43 ± 0.97	0.029 ± 0.007	0.015 ± 0.009	26	*qSLW-X3*	0.0	2.51–4.13	−0.010 to −0.007	0.002–0.014	4.62–9.02	JDRC ($F_{2:4}$), DC($F_{2:4}$),RC($F_{2:4}$)

Table 4 continued

Trait	Chr	Marker interval	Cytoplasm-by-QTL interaction						Composite interval mapping						
			QTL[a]	Position (cM)	LOD ± SD	Additive ± SD	Dominant ± SD	Ratio (%)	QTL	Position (cM)	LOD	Additive	Dominant	PVE(%)	Population source
SLT	D1b-2, 2	satt266-satt282	*qSLT-2c*	34.8	3.82 ± 0.70	0.033 ± 0.009	−0.096 ± 0.009	36	*qSLT-2-2*	34.9	2.50	−0.014	0.019	4.31	$RC(F_{2:5})$
	C2-2, 6	satt277-satt643	*qSLT-6c*	38.3	3.80 ± 1.04	−0.014 ± 0.008	0.100 ± 0.013	78	*qSLT-6*	34.4	2.77	−0.022	0.017	6.90	$DC(F_{2:4})$
	J, 16	satt132-satt414	*qSLT-16c-2*	12.5	3.42 ± 0.91	−0.069 ± 0.009	0.000 ± 0.016	42	*qSLT-16-1*	14.0	3.24	0.016	0.001	6.17	$RC(F_{2:5})$
SWT	J, 16	sat_165-satt132	*qSWT-16c*	11.4	5.72 ± 1.39	−0.053 ± 0.006	−0.010 ± 0.012	90	*qSWT-16*	9.0	3.99	−0.010	−0.002	9.55	$DC(F_{2:5})$

[a] QTL nomenclature is similar to that for main-effect QTL except for following a letter *c* after chromosome number

Table 5 Environment-by-QTL interactions that are consistent with main-effect QTL detected by composite interval mapping

Trait	Chr	Marker interval	Environment-by-QTL interaction						Composite interval mapping						
			QTL[a]	Position	LOD ± SD	Additive ± SD	Dominant ± SD	Ratio (%)	QTL	Position (cM)	LOD	Additive	Dominant	PVE (%)	Population source
SW	C2-2, 6	satt643-staga001	*qSW-6e*	52.4	3.54 ± 0.67	0.040 ± 0.004	0.006 ± 0.010	48	*qSW-6-2*	59.0	3.41	−0.031	−0.095	8.63	Direct cross (DC)($F_{2:4}$)
						−0.014 ± 0.006	−0.008 ± 0.010								
						−0.026 ± 0.004	0.001 ± 0.012								
	O-2, 10	satt592-sat_274	*qSW-10e*	46.1	2.51 ± 0.24	0.014 ± 0.006	0.034 ± 0.014	10	*qSW-10-2*	18.0	3.12	−0.025	−0.089	7.66	Reciprocal cross (RC)$F_{2:5}$)
						−0.003 ± 0.003	0.006 ± 0.018								
						−0.011 ± 0.004	−0.040 ± 0.004								
ST	C2-1, 6	satt640-satt422	*qST-6e*	23.3	3.63 ± 1.18	0.026 ± 0.006	−0.006 ± 0.004	22	*qST-6-1*	15.0	3.72	−0.084	−0.014	7.84	$RC(F_{2:5})$
						0.007 ± 0.005	−0.027 ± 0.006								
						−0.032 ± 0.004	0.033 ± 0.008								
SWT	C2-1, 6	satt640-satt422	*qSWT-6e*	25.5	2.93 ± 0.41	−0.042 ± 0.005	0.003 ± 0.011	26	*qSWT-6*	25.0	3.97	−0.015	0.004	6.71	$RC(F_{2:3})$
						0.000 ± 0.004	0.016 ± 0.007								
						0.041 ± 0.004	−0.019 ± 0.009								

[a] QTL nomenclature is similar to that for main-effect QTL except for following a letter *e* after chromosome number

We tried to compare the QTL in the present study with those in previous studies conducted by other research groups. Although it seemed difficult because a few common markers existed between their data and ours, it was rewarding to find that some loci that we detected were also detected by other researchers, for example, four: *qSL-19*, *qSW-19-2*, *qST-19* and *qST-7-1*, one: *qSL-7-2* and four QTL: *qST-10-1*, *qSLW-10-1*, *qSLT-10-1* and *qSLW-18* in this study were also detected by Salas et al. (2006), Li et al. (2008) and Liang et al. (2008), respectively. One nucleocytoplasmic interaction, *qST-10c-2*, associated with marker satt331, had the same position as the M-QTL *qST_O_1a* detected by Liang et al. (2008). In addition, some QTL detected in the present study are potentially similar to those in previous studies because they are located on the same chromosome; for example, some QTL identified on chromosomes 2, 3, 4, 5, 6, 7, 13, 15 and 19 by Salas et al. (2008), 7, 13 and 16 by Li et al. (2008) and 1, 5, 10, 13, 14, 16, 17 and 20 by Liang et al. (2008).

The phenomenon of QTL clusters has previously been reported in rice (Cai and Morishima 2002), wheat (Quarrie et al. 2006), cotton (Shappley et al. 1998; Qin et al. 2008), rat (Stoll et al. 2000) and sorghum (Lin et al. 1995). Previous work has indicated that numerous disease resistance loci are clustered in various regions of the soybean genome; for example, chromosomes D1b and F (Rector et al. 1999; Hayes et al. 2000). This phenomenon was also evident in our results. Although the common QTL for seed size and shape traits were distributed on 16 chromosomes, the majority of the QTL were clustered in eight chromosomal intervals (Fig. 1). The eight intervals, located on chromosomes 3, 5, 6, 10, 13, 17, 18 and 20, were found to be involved in the control of two or more of the above traits. The QTL for SL, SW, ST and SWT were clustered in the satt640-satt422 interval on chromosome 6, shared the same direction of both additive and dominant effects and involved novel alleles from the same parent, N493. The QTL for SW and ST were located in the satt449-sat_356 interval on chromosome 5, shared the same direction of both additive and dominant effects and involved novel alleles from the same parent, N493, as well. The QTL for ST and SLW were located in the satt449-sat_356 interval on chromosome 5 and exhibited different directions of both additive and dominant effects, which may have caused the significant negative correlation between ST and SLW. The QTL for SL, SW, SLW and SLT were clustered between satt331 and satt592 and between satt592 and sat_274 on chromosome 10, and they exhibited the same phenomenon found by Liang et al. (2008) that some QTL were clustered between satt331 and sat_038. The QTL for SLW, SLT and SWT were clustered in the satt649-satt269 interval on chromosome 13, and the QTL for SL, SW, ST and SLW were clustered in the satt309-satt688 interval on chromosome 18. The cause of QTL clustering has been discussed by Qin et al. (2008). Considering the high correlation between some traits, coupling tight linkage and pleiotropy could better explain the data (Qin et al. 2008). In addition, Xiong et al. (1999) and Bres-Patr et al. (2001) performed a QTL analysis of domestication-related traits using crosses between *O. sativa* and *O. rufipogon* and between *O. sativa* and weedy-type rice. Both studies demonstrated clusters of QTL related to domestication. In the present study, the evidence of QTL in soybean related to domestication is not shown but needs to be further studied in the future.

To date, there have been some reports about the mechanism of seed development, but most of them concentrate on the model plants *Arabidopsis* and rice (Sundaresan 2005; Song et al. 2007; Shomura et al. 2008). In rice, the loss of *GW2* function increases cell number, resulting in a larger (wider) spikelet hull, and it accelerates the grain milk filling rate, resulting in enhanced grain width, weight and yield (Song et al. 2007). The deletion in *qSW5* in rice affects seed width and results in a significant increase in sink size owing to an increase in cell number in the outer glume of the rice flower (Shomura et al. 2008). In soybean, there are no reports on this topic. This also needs to be studied in the future.

Acknowledgments We are grateful to the Associate Editor Dr. Rajcan, Istvan, and two anonymous reviewers for their constructive comments and suggestions that significantly improved the manuscript. The work was supported by the 973 program (2011CB109300), Jiangsu Natural Science Foundation (BK2008335), the National Natural Science Foundation of China (30971848), the 111 Project (B08025) and NCET (NCET-05-0489) to YMZ.

References

Ayoub M, Symons SJ, Edney MJ, Mather DE (2002) QTLs affecting kernel size and shape in a two-rowed by six-rowed barley cross. Theor Appl Genet 105:237–247

Bres-Patr YC, Lorieux M, Clement G, Bangratz M, Ghesquiere A (2001) Heredity and genetic mapping of domestication-related traits in a temperate *Japonica* weedy rice. Theor Appl Genet 102:118–126

Cai HW, Morishima H (2002) QTL clusters reflect character associations in wild and cultivated rice. Theor Appl Genet 104:1217–1228

Cober ER, Voldeng HD, Fregeau-Reid JA (1997) Heritability of seed shape and seed size in soybean. Crop Sci 37:1767–1769

Cui CX, Xuan YN (2007) Analysis of the factors affecting soybean trade in China and suggestions of strategies. World Agric 29:7–10

Dou BD, Hou BW, Wang F, Yang JB, Ni ZF, Sun QX, Zhang YM (2010) Further mapping of quantitative trait loci for female sterility in wheat (*Triticum aestivum* L.). Genet Res 92:63–70

Fan CC, Xing YZ, Mao HL, Lu TT, Han B, Xu CG, Li XH, Zhang QF (2006) *GS3*, a major QTL for grain length and weight and minor

QTL for grain width and thickness in rice, encodes a putative transmembrane protein. Theor Appl Genet 112:1164–1171

Gegas VC, Nazari A, Griffiths S, Simmonds J, Fish L, Orford S, Sayers L, Doonan LH, Snape JW (2010) A genetic framework for grain size and shape variation in wheat. Plant Cell 22:1046–1056

Hayes AJ, Ma GR, Buss GR, Saghai Maroof MA (2000) Molecular marker mapping of *Rsv*4, a gene conferring resistance to all known strains of soybean mosaic virus. Crop Sci 40: 1434–1437

Johnson SL, Fehr WR, Welke GA, Cianzo SR (2001) Genetic variability for seed size of two- and three-parent soybean populations. Crop Sci 41:1029–1033

Kazuyoshi T, Ayumi Y (1980) Major gene for controlling grain size in rice. Breed J 30:280–282

Kim H, Lee S, Park K, Lee Y (2000) Identification of quantitative trait loci associated with seed size and weight in soybean. Korean J Crop Sci 45:227–231

Kosambi DD (1944) The estimation of map distance from recombination values. Ann Eugen 12:172–175

Lander ES, Kruglyak L (1995) Genetic dissection of complex traits guidelines for interpreting and reporting linkage results. Nat Genet 11:241–247

Leroy AR, Cianzio SR, Fehr WR (1991) Direct and indirect selection for small seed of soybean in temperate and tropical environments. Crop Sci 31:697–699

Li CD, Jiang HW, Zhang WB, Qiu PC, Liu CY, Li WF, Gao YL, Chen QS, Hu GH (2008) QTL analysis of seed and pod traits in soybean. Mol Plant Breed 6:1091–1100

Liang HZ, LI WD, Wang H, Fang XJ (2005) Genetic effects on seed traits in soybean. Acta Genetica Sinica 32:1199–1204

Liang HZ, Wang SF, Yu YL, Wang TF, Gong PT, Fang XJ, Liu XY, Zhao SJ, Zhang MC, Li WD (2008) Mapping quantitative trait loci for six seed shape traits in soybean. Henan Agric Sci 45:54–60

Lin YR, Schertz KF, Paterson AH (1995) Comparative analysis of QTL affecting plant height and maturity across the Poaceae, in reference to an interspecific sorghum population. Genetics 141:391–411

Lipp M, Brodmann P, Pietsch K, Pauwels J, Anklam E (1999) IUPAC collaborative trail study of a method to detect genetically modified soybeans and maize in dried powder. J AOAC Int 82:923–928

Mansur LM, Orf JH, Lark KG (1993) Determining the linkage of quantitative trait loci to RFLP marker using extreme phenotypes of recombinant inbreds of soybean (*Glycine max* L. Merr.). Theor Appl Genet 86:914–918

Mansur LM, Org 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

McCouch SR, Chen X, Panaud O, Temnykh S, Xu Y, Cho YG, Huang N, Ishiii T, Blair M (1997) Microsatellite marker development, mapping and applications in rice genetics and breeding. Plant Mol Biol 35:89–99

Nelson RL, Wang P (1989) Variation and evaluation of seed shape in soybean. Crop Sci 29:147–150

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

Qin HD, Guo WZ, Zhang YM, Zhang TZ (2008) QTL mapping of yield and fiber traits based on a four-way cross population in *Gossypium hirsutum* L. Theor Appl Genet 117:883–894

Quarrie SA, Quarrie SP, Radosevic R, Rancic D, Kaminska A, Barnes JD, Leverington M, Ceoloni C, Dodig D (2006) Dissecting a wheat QTL for yield present in a range of environments: from the QTL to candidate genes. J Exp Bot 57:2627–2637

Rabiei B, Valizadeh M, Ghareyazie B, Moghaddam M, Ali AJ (2004) Identification of QTLs for rice grain size and shape of Iranian cultivars using SSR markers. Euphytica 137:325–332

Rector BG, All JN, Parrott WA, Boerma HR (1999) Quantitative trait loci for antixenosis resistance to corn earworm in soybean. Crop Sci 39:531–538

Sakamoto T, Matsuoka M (2008) Identifying and exploiting grain yield genes in rice. Curr Opin Plant Biol 11:209–214

Salas P, Oyarzo-Llaipen JC, Wang D, Chase K, Mansur L (2006) Genetic mapping of seed shape in three population of recombinant inbred lines of soybean (*Glycine max* L. Merr.). Theor Appl Genet 113:1459–1466

Santos FR, Pena SDJ, Epplen JT (1993) Genetic and population study of an Y-linked tetranucleotide repeat DNA polymorphism with a simple non-isotopic technique. Hum Genet 90:655–656

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183

Sen S, Churchill GA (2001) A statistical framework for quantitative trait mapping. Genetics 159:371–387

Shappley ZW, Jenkins JN, Zhu J, McCarty JC (1998) Quantitative trait loci associated with yield and fiber traits of upland cotton. J Cotton Sci 4:153–163

Shi CH, Shen ZT (1994) Analysis of genetic effects of grain traits in indica rice. J Zhejiang Agric Univ 20:405–410

Shomura A, Izawa T, Ebana K, Ebitani T, Kanegae H, Konishi S, Yano M (2008) Deletion in a gene associated with grain size increased yields during rice domestication. Nat Genet 40: 1023–1028

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

Song XJ, Huang W, Shi M, Zhu MZ, Lin HX (2007) A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. Nat Genet 39:623–630

Stoll M, Kwitek-Black AE, Cowley AW, Harris EL, Harrap SB, Krieger JE, Printz MP, Provoost AP, Sassard J, Jacob HJ (2000) New target regions for human hypertension via comparative genomics. Genome Res 10:473–482

Sundaresan V (2005) Control of seed size in plants. Proc Natl Acad Sci USA 102:17887–17888

Van Ooijen JW, (2006) JoinMap Version 4.0: software for the calculation of genetic linkage in experimental populations. CPRO–DLO, Wageningen

Voorrips RE (2002) MapChart: Software for the graphical presentation of linkage maps and QTLs. J Hered 93:77–78

Wan XY, Weng JF, Zhai HQ, Wang JK, Lei CL, Liu XL, Guo T, Jiang L, Su N, Wan JM (2008) Quantitative trait loci (QTL) analysis for rice grain width and fine mapping of an identified QTL allele *gw-5* in a recombination hotspot region on chromosome 5. Genetics 179:2239–2252

Wang S, Basten CJ, Zeng ZB (2007) Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh

Wilson DO Jr (1995) Storage of orthodox seeds. In: Basra AS (ed) Seed quality: basic mechanisms agricultural implications. Food Products Press, New York, pp 173–208

Xiong LZ, Liu KD, Dai XK, Xu CG, Zhang Q (1999) Identification of genetic factors controlling domestication-related traits of rice using an F_2 population of a cross between *Oryza sativa* and *O* rufipogon. Theor Appl Genet 98:243–251

Xu SZ, Jia ZY (2007) Genomewide analysis of epistatic effect for quantitative trait in barley. Genetics 175:1955–1963

Zhang YM, Xu S (2005) A penalized maximum likelihood method for estimating epistatic effects of QTL. Heredity 95:96–104

Zheng YL, Lai ZM, Yang KC (1985) The relationship between maize seed traits and seed size and the study on the genetics. J Sichuan Agric Univ 8:73–79