2014

Identification and validation of quantitative trait loci for seed yield, oil and protein contents in two recombinant inbred line populations of soybean

Xianzhi Wang

South Dakota State University

Guo-Liang Jiang

South Dakota State University

Marci Green

South Dakota State University

Roy A. Scott

Crop Production and Protection, GWCC-BLTSVL, USDA/ARS, Beltsville, MD

Qijian Song

Soybean Genomics and Improvement Laboratory, USDA/ARS, Beltsville, MD, qijian.song@ars.usda.gov

See next page for additional authors

Follow this and additional works at: <https://digitalcommons.unl.edu/agronomyfacpub>

Part of the [Agricultural Science Commons](#), [Agriculture Commons](#), [Agronomy and Crop Sciences Commons](#), [Botany Commons](#), [Horticulture Commons](#), [Other Plant Sciences Commons](#), and the [Plant Biology Commons](#)

Wang, Xianzhi; Jiang, Guo-Liang; Green, Marci; Scott, Roy A.; Song, Qijian; Hyten, D. L.; and Cregan, P. B., "Identification and validation of quantitative trait loci for seed yield, oil and protein contents in two recombinant inbred line populations of soybean" (2014). *Agronomy & Horticulture -- Faculty Publications*. 814.
<https://digitalcommons.unl.edu/agronomyfacpub/814>

This Article is brought to you for free and open access by the Agronomy and Horticulture Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Agronomy & Horticulture -- Faculty Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Authors

Xianzhi Wang, Guo-Liang Jiang, Marci Green, Roy A. Scott, Qijian Song, D. L. Hyten, and P. B. Cregan

Identification and validation of quantitative trait loci for seed yield, oil and protein contents in two recombinant inbred line populations of soybean

Xianzhi Wang · Guo-Liang Jiang · Marci Green ·
Roy A. Scott · Qijian Song · David L. Hyten ·
Perry B. Cregan

Received: 19 March 2014 / Accepted: 6 May 2014 / Published online: 27 May 2014
© Springer-Verlag Berlin Heidelberg 2014

Abstract Soybean seeds contain high levels of oil and protein, and are the important sources of vegetable oil and plant protein for human consumption and livestock feed. Increased seed yield, oil and protein contents are the main objectives of soybean breeding. The objectives of this study were to identify and validate quantitative trait loci (QTLs) associated with seed yield, oil and protein contents in two recombinant inbred line populations, and to evaluate the consistency of QTLs across different environments, studies and genetic backgrounds. Both the mapping population (SD02-4-59 × A02-381100) and validation population (SD02-911 × SD00-1501) were phenotyped for the three traits in multiple environments. Genetic analysis indicated that oil and protein contents showed high heritabilities while yield

exhibited a lower heritability in both populations. Based on a linkage map constructed previously with the mapping population and using composite interval mapping and/or interval mapping analysis, 12 QTLs for seed yield, 16 QTLs for oil content and 11 QTLs for protein content were consistently detected in multiple environments and/or the average data over all environments. Of the QTLs detected in the mapping population, five QTLs for seed yield, eight QTLs for oil content and five QTLs for protein content were confirmed in the validation population by single marker analysis in at least one environment and the average data and by ANOVA over all environments. Eight of these validated QTLs were newly identified. Compared with the other studies, seven QTLs for seed yield, eight QTLs for oil content and nine QTLs for protein content further verified the previously reported QTLs. These QTLs will be useful for breeding higher yield and better quality cultivars, and help effectively and efficiently improve yield potential and nutritional quality in soybean.

Communicated by S. Hohmann.

X. Wang · G.-L. Jiang (✉) · M. Green
Plant Science Department, South Dakota State University,
North Plains Biostress Laboratory Building, Box 2140C,
Brookings, SD 57006, USA
e-mail: guo-liang.jiang@sdstate.edu; gljiang99@yahoo.com

G.-L. Jiang
Agricultural Research Station, Virginia State University,
P.O. Box 9061, Petersburg, VA 23806, USA

R. A. Scott
Crop Production and Protection, GWCC-BLTSLV, USDA/ARS,
5601 Sunnyside Ave, Beltsville, MD 20705-5139, USA

Q. Song · D. L. Hyten · P. B. Cregan
Soybean Genomics and Improvement Laboratory, USDA/ARS,
10300 Baltimore Ave, Beltsville, MD 20705-2350, USA

Present Address:

D. L. Hyten
Pioneer Hi-Bred International, 7300 NW 62nd Ave, Po Box 1004,
Johnston, IA 50131, USA

Keywords Soybean · Yield · Oil · Protein · Quantitative trait loci (QTLs) · Molecular breeding

Introduction

Soybean (*Glycine max* (L.) Merr.) is one of the major field crops grown worldwide. Seeds of soybean are rich in oil (averaged 20 %) and protein (averaged 40 %) (Clemente and Cahoon 2009), and thus soybean is an important source of vegetable oil for human consumption and industrial applications and is also an important source of plant protein for human food and livestock feed (Chiari et al. 2004; Yesudas et al. 2013). Increases in seed oil and protein contents of soybean would enhance the competitiveness of the crop. It has been reported that within the

USDA soybean germplasm collection, the phenotypic variation ranged from 8.1 to 27.9 % for oil and from 34.1 to 56.8 % for protein, respectively (Wilson 2004). It indicates that there is a great potential for the improvement of soybean oil and protein. However, simultaneous increases in both oil and protein contents can be realized to a limited extent only, since there is generally a significant negative correlation between oil and protein contents in soybean seeds (Burton 1987). Improvement of the overall yield of soybean means more oil and protein in terms of production per unit area. Therefore, increasing seed yield potential is the most important objective of soybean breeding. Identification and validation of QTLs associated with soybean yield, oil and protein contents will help the improvement of these important traits.

Soybean seed yield, oil and protein contents are quantitatively inherited traits controlled by multiple genes that may show small or large effects. Numerous QTLs for seed yield, oil and protein contents have been previously identified in soybean and each of the 20 chromosomes (Chr) carries one or more QTLs (Diers et al. 1992; Lee et al. 1996; Brummer et al. 1997; Orf et al. 1999; Csanadi et al. 2001; Yuan et al. 2002; Wang et al. 2004; Hyten et al. 2004; Guzman et al. 2007; Palomeque et al. 2009a, b; Yesudas et al. 2013). Diers et al. (1992) mapped two major QTLs controlling oil and protein contents on linkage groups E and I, which were originally named linkage groups A and K by Diers et al. (1992) (Sebolt et al. 2000). Brummer et al. (1997) evaluated eight different populations from the Midwest USA for seed oil and protein contents in multiple environments. Seven QTLs for oil content and nine QTLs for protein content were detected in one or more populations. Hyten et al. (2004) identified four QTLs for protein content, six QTLs for oil content and seven QTLs for seed size in a recombinant inbred line (RIL) population consisting of 131 F_6 -derived lines. Kabelka et al. (2004) used the population of BSR 101 $\times$ LG82-8379 to map QTLs associated with seed yield and other agronomic traits in 12 environments. Fifteen QTLs for seed yield, 3 QTLs for oil content and 11 QTLs for protein content were detected. Wang et al. (2004) identified four seed yield QTLs across environments on linkage groups C2, E, K and M in five populations each consisting of 57–112 BC_2F_4 -derived lines from a cross of *Glycine max* $\times$ *Glycine soja*. Guzman et al. (2007) evaluated the agronomic traits and yield performance of three backcross-derived populations in 2 years. Thirteen QTLs for seed yield were mapped on linkage groups A1, B2, C1, C2, J, K, L and O. More recently, Yesudas et al. (2013) detected 11 QTLs for oil and protein contents and seed weight on linkage groups A1, A2, B1, C2, D1b, E, H, I and N in the RIL population EF94.

One of the objectives of QTL analysis in plants is to facilitate the application of molecular markers to practical

breeding, i.e. marker-assisted selection (MAS). For MAS, favorable alleles could be easily introgressed into and then expressed in an elite line if they are independent of environments and genetic backgrounds (Palomeque et al. 2010). Although numerous QTLs for seed yield, oil and protein contents have been identified in soybean, the QTLs that can be consistently verified across multiple environments and different genetic backgrounds are still very limited. To date, only a few QTLs for seed yield, oil and protein contents have been verified. Orf et al. (1999) used amplified fragment length polymorphism (AFLP) markers and simple sequence repeat (SSR) markers to identify QTLs for agronomic traits across four environments in three RIL populations. Five QTLs for seed yield, six QTLs for oil content and five QTLs for protein content were detected, but most of them were detected in only one population and no QTL could be confirmed in all three populations. Csanadi et al. (2001) identified four markers for protein content, three markers for oil content and eight markers for seed weight in an F_2 population derived from a cross of two early maturing soybean cultivars. Only 4 of the 15 QTL regions were also reported in previous studies (Csanadi et al. 2001). Concibido et al. (2003) mapped a QTL for seed yield on linkage group B2 from a population of 265 BC_2 individuals derived from a cross between HS-1 and wild soybean (*Glycine soja* Sieb. and Zucc.) PI 407305. They further assessed the QTL effects in various elite soybean genetic backgrounds, and found that the effect of the yield QTL was observed in only 2 of 6 genetic backgrounds. Li et al. (2008) used two BC_2F_4 populations to map and validate QTLs for yield and yield components in three environments. Eleven QTLs were mapped but only one QTL for seed yield, which was linked to the marker Satt511 on linkage group A1, was confirmed in the two populations. Palomeque et al. (2009a) mapped five mega-environment-universal and two mega-environment-specific QTLs for seed yield in a RIL population derived from the cross OAC Millennium $\times$ Heinong 38 across multiple environments in China and Canada. But none of the seven QTLs for seed yield was confirmed in the validation RIL population of Pioneer 9071 $\times$ 8902 in their subsequent study (Palomeque et al. 2010). Therefore, validation of known QTLs and identification of new QTLs will facilitate the effective use of MAS in practical breeding.

The objectives of this study were: (1) to identify QTLs for seed yield, oil and protein contents in a RIL mapping population of SD02-4-59 $\times$ A02-381100 in five environments; (2) to validate the QTLs using a different RIL population derived from a cross of SD02-911 $\times$ SD00-1501 in three environments; and (3) to confirm previously identified QTLs and determine the stable QTLs which were consistent with other studies.

Materials and methods

Plant materials and trait evaluation

Two RIL populations of soybean were used for QTL mapping and validation in this study. The mapping population for QTL identification of seed yield, oil and protein contents consisted of 87 F_5 -derived RILs by single-seed decent (SSD) from the cross of SD02-4-59 $\times$ A02-381100, as described previously in QTL analysis of fatty acids (Wang et al. 2012, 2014). The parent SD02-4-59 was a breeding line developed by South Dakota State University, and A02-381100 was a low-linolenic line developed by Iowa State University. The validation population, consisting of 196 F_5 -derived lines, was developed by single-pod decent (a modified SSD) (Fehr 1987) from a cross of SD02-911 $\times$ SD00-1501. Both of the parents were developed by South Dakota State University, SD02-911 was a breeding line and SD00-1501 is a released high-protein line (PI 662943).

The mapping population was phenotyped for seed yield, oil and protein contents in five environments. All the 87 F_5 -derived RILs were planted at Aurora, SD in 2009 and 2010 (designated as E09AU and E10AU, respectively), Beresford, SD in 2009 (E09BF) and 2011 (E11BF), and Volga, SD in 2011 (E11VG). Due to lack or loss of seeds, the parents were not included in the phenotyping experiments (Wang et al. 2012, 2014). In order to confirm the presence and effects of the QTLs identified in the mapping population, the validation population was evaluated for seed yield, oil and protein contents in three environments, i.e. all 196 F_5 -derived RILs and two parents were planted at Aurora, SD in 2011 (E11AU), and Volga, SD (E12VG) and Brookings, SD (E12BK) in 2012. Both the mapping and validation populations were planted in the same plot technique and experimental design in all environments, i.e. the RILs were planted in two-row plots with two replications in a randomized complete block design (RCBD). Plots were planted at 26 seeds m^{-1} in rows 4.42 m long with a 0.76 m row spacing. After full maturity (R8), all plots were harvested with a plot combine and the yield data ($kg\ ha^{-1}$) was obtained on a 13 % moisture basis. All seeds from each plot were dried in an air-dryer, and then sieved and cleaned by hand to remove the remnants of pods and branches and the broken seeds. Cleaned seeds were stored in cardboard boxes for the seed composition analysis. Seed oil and protein contents were determined using a DA7200 near-infrared (NIR) analyzer (Perten Instruments, Sweden). According to the manufacturer's guidelines, approximately 290 g of cleaned whole seeds were loaded into the sample cup and data were read three times per sample and the averages were used in statistical analysis. Data of seed oil and protein contents were presented on a 0 % moisture basis.

Statistical analysis

ANOVA was performed to determine the significance of genotypic differences among the RILs and genotype by environment interactions in the mapping and validation populations using SAS version 9.1 (SAS Institute Inc., Cary, NC). Heritability was estimated on the line mean basis as $h^2 = \sigma_g^2 / [\sigma_g^2 + (\sigma_{ge}^2/e) + (\sigma^2/re)]$, where h^2 represents the heritability, σ_g^2 is genotypic variance, σ_{ge}^2 is genotype $\times$ environment interaction variance, σ^2 is error variance, r is number of replications and e is number of environments (Fehr 1987).

Linkage map and QTL analysis

The linkage map used for the mapping population in this study was the same as that was previously used in QTL analysis of unsaturated fatty acids (Wang et al. 2014). A total of 1,428 single nucleotide polymorphism (SNP) and 1,077 SSR markers were screened for polymorphism in the mapping population. Of them, 516 SNP and 477 SSR markers exhibited polymorphism and, as well as three *GmFAD3* genes, were used to genotype the population. The markers for which data were missing in more than 10 lines and which exhibited significant segregation distortion (i.e. significant at $P = 0.01$) were excluded from the map construction (Wang et al. 2014). Except those markers that could not be mapped to any linkage group, a total of 311 SNP and 399 SSR markers as well as the three *GmFAD3* genes were finally mapped on a linkage map, which covered all 20 soybean chromosomes and spanned a total length of 2,099.9 cM with an average interval length of 3.2 cM (Wang et al. 2014).

QTL analysis was conducted in WinQTLCart version 2.5 (Wang et al. 2005) and QTL IciMapping version 3.1 (Wang et al. 2011) for each environment and the average data over five environments in the mapping population. Single marker analysis (SMA), interval mapping (IM), and composite interval mapping (CIM) were performed. Based on permutation tests performed 1,000 times at $\alpha = 0.05$ for experiment-wise Type I error and referring to the empirical threshold values widely used for QTL mapping (Bachlava et al. 2009; Panthee et al. 2006; Tucker et al. 2010), a LOD value of 2.5 was set as the threshold for significance of a QTL. In a few cases, the QTLs with a LOD value above 2.0 (equivalently $P = 0.002$ and significant at $P < 0.01$ for ANOVA) were also declared significant to avoid missing of QTLs due to slightly lower significance (Cornelius et al. 2005; Wang et al. 2014; Yesudas et al. 2013). However, the QTLs that could be detected in only one environment in the mapping population were not declared significant in this paper. Therefore, only the QTLs that were detected by CIM and/or IM as well as SMA in at least two individual

Table 1 Descriptive statistics for yield (kg ha^{-1}), oil and protein contents (%) in the mapping population of SD02-4-59 $\times$ A02-381100 over five environments (2009–2011) and in the validation population of SD02-911 $\times$ SD00-1501 over three environments (2011 and 2012)

Trait	Mean	Range	F_g	F_e	$F_{g \times e}$	$\text{LSD}_{0.05}$	h^2
Mapping population							
Yield	$2,285.1 \pm 241.8$	1,784.9–3,065.7	6.2**	957.9**	2.2**	269.8	0.65
Oil	18.4 ± 0.9	16.3–20.4	71.0**	939.7**	4.4**	0.3	0.94
Protein	42.4 ± 1.3	40.2–46.0	71.6**	1,097.6**	4.4**	0.4	0.94
Validation population							
Yield	$2,477.6 \pm 170.3$	1,748.1–3,012.5	1.9**	308.4**	1.6**	325.3	0.19
Oil	20.9 ± 0.6	17.3–21.8	11.2**	319.9*	1.9**	0.5	0.84
Protein	39.7 ± 0.8	38.3–43.6	8.1**	279.0**	2.4**	0.8	0.71

F_g , F_e , and $F_{g \times e}$ represent F values for genotype, environment and genotype $\times$ environment interaction, respectively

h^2 heritability

* and ** significant at $P < 0.05$ and 0.01 , respectively

environments or in one environment and the average data over all five environments are presented in this paper. To verify the independence of the QTLs/loci which were located on the same linkage group with the same source of favorable alleles, a comparison of two-locus combinations was performed based on the results of ANOVA (Jiang et al. 2007).

QTL validation

To validate the QTLs identified in the mapping population, all the 1,077 SSR markers screened in the mapping population were also screened for polymorphism between SD02-911 and SD00-1501, the parents of the validation population. All the polymorphic SSR markers on the linkage groups that carried QTLs for seed yield, oil and/or protein content in the mapping population were selected to genotype all the RILs of the validation population. SMA was performed in WinQTLCart version 2.5 (Wang et al. 2005) and QTL IciMapping version 3.1 (Wang et al. 2011) to detect the associations between the markers and the traits of interest. As described above, a LOD value of 2.5 was set as the threshold for significance of a QTL in the validation population. A comparison between two groups of RILs carrying different marker alleles was also conducted based on ANOVA over all environments to verify the significance of the marker/locus-trait associations (Jiang et al. 2007).

Results

Phenotypic analysis and heritability

ANOVA results showed that the differences among RILs in both mapping and validation populations were highly significant for all three traits ($P < 0.01$, Table 1). The

environmental differences and genotype $\times$ environment interaction effects were also highly significant for both populations ($P < 0.01$, Table 1). Over all environments, the mapping population exhibited a higher average protein content but lower averages of seed yield and oil content than the validation population (Table 1), and the differences between extremes were similar in the two populations though they were not evaluated simultaneously. The range of variation for each of the traits was quite large for both populations (Table 1). For the validation population, transgressive segregation was observed for all three traits (the averages of SD02-911 and SD00-1501 were 2757.3 and 2,125.1 kg ha^{-1} for yield, 21.7 and 20.0 % for oil content, and 38.8 and 43.2 % for protein content, respectively). Of the three traits investigated, oil content showed high heritability in both populations (0.94 and 0.84, respectively). Yield showed moderate heritability in the mapping population (0.65), but low heritability (0.19) in the validation population. For protein content, the estimate of heritability was high (0.94) in the mapping population and moderate (0.71) in the validation population (Table 1).

QTL analysis in the mapping population

A total of 39 QTLs were identified for seed yield, oil and protein contents in the mapping population. For seed yield, 12 QTLs were mapped on linkage groups B2, D2, E, F, G, I, J, K, M and O by CIM and/or IM analysis (Table 2). On both linkage group D2 and I, two QTLs with the same source of favorable alleles for seed yield were detected. The comparison of QTL combinations based on ANOVA showed that there were cumulative effects for the two QTLs on the same linkage group (Table 3), indicating that the two QTLs on either linkage group D2 or I were independent of each other. The QTLs on linkage groups B2, D2 and I ($qYIE-B2$, $qYIE-D2-2$ and $qYIE-I-2$) were repeatedly

Table 2 QTLs for seed yield, oil and protein contents detected by CIM/IM method in the mapping population of SD02-4-59 × A02-381100

QTL ^a	LG (Chr #)	Peak marker ^b	Interval		LOD	A.E ^d	R ² (%)	Env ^e	References ^f
			Int-L ^c	Int-R ^c					
Yield									
<i>qYIE-B2</i>	B2 (14)	BARCSOYSSR_14_1375	BARC-059265-15700	BARCSOYSSR_14_1375	3.0–3.4	126.4 to 195.7	14.4 to 20.2	1–3, 0	VAL; 4, 6, 12, 18
<i>qYIE-D2-1</i>	D2 (17)	Set_192	BACR0YSSR_17_0118	BARC-021477-04128	3.3–6.3	65.2 to 222.6	6.9 to 20.6	4, 0	VAL
<i>qYIE-D2-2</i>	D2 (17)	BARCSOYSSR_17_0900	BARCSOYSSR_17_0844	BARC-056215-14153	2.2–5.3	111.0 to 169.5	10.9 to 24.0	1–3, 0	18
<i>qYIE-E</i>	E (15)	Satt685	BARCSOYSSR_15_1453	Satt685	3.0–3.2	92.8 to 126.4	14.6 to 16.0	3, 0	
<i>qYIE-F</i>	F (13)	BARCSOYSSR_13_0105	BARCSOYSSR_13_0119	BARCSOYSSR_13_0105	2.2–3.8	–100.2 to –221.3	11.1 to 18.2	1, 2, 0	VAL
<i>qYIE-G</i>	G (18)	BARCSOYSSR_18_0306	Satt235	BARCSOYSSR_18_0306	2.3–2.7	123.1 to 184.9	11.3 to 13.2	2, 3	
<i>qYIE-I-1</i>	I (20)	BARCSOYSSR_20_0049	BARCSOYSSR_20_0049	BARCSOYSSR_20_0079	2.7–4.9	111.6 to 231.3	19.0 to 30.6	1, 2, 0	
<i>qYIE-I-2</i>	I (20)	BARCSOYSSR_20_0750	BARC-041129-07912	BARCSOYSSR_20_0782	2.4–8.9	119.7 to 207.1	15.9 to 37.3	1–3, 0	18
<i>qYIE-J</i>	J (16)	Satt547	Satt_370	Satt547	2.5–8.5	88.1 to 129.1	6.7 to 27.8	3, 0	6
<i>qYIE-K</i>	K (9)	Satt273	Satt273	BARCSOYSSR_09_1171	2.5–2.8	–92.1 to –188.3	13.4 to 14.5	1, 2, 0	6, 9, 18, 19, 20, 21
<i>qYIE-M</i>	M (7)	Satt540	Satt540	BARC-023593-05477	2.2–3.3	114.3 to 182.2	14.3 to 21.1	4, 0	VAL; 9, 12, 18, 19, 20
<i>qYIE-O</i>	O (10)	BARC-050013-09288	BARCSOYSSR_10_1395	BARC-050013-09288	2.6–8.6	69.3 to 160.7	5.6 to 24.0	1–3	VAL; 18
Oil									
<i>qOIL-A1</i>	A1 (5)	BARC-060051-16321	Sat_374	BARC-029787-06340	2.9–7.6	0.3 to 0.5	6.6 to 23.0	1, 5, 0	1, 12, 19
<i>qOIL-B1</i>	B1 (11)	BARCSOYSSR_11_0545	BARC-016539-02087	BARCSOYSSR_11_0545	2.1–2.8	–0.3 to –0.5	12.7 to 16.2	1, 2, 4, 0	14
<i>qOIL-C1-1</i>	C1 (4)	Sat_140	Sat_140	BARCSOYSSR_04_0366	2.2–2.4	–0.3 to –0.4	13.2 to 15.0	2, 3, 5, 0	VAL
<i>qOIL-C1-2</i>	C1 (4)	BARC-044521-08714	BARCSOYSSR_04_0343	BARC-040777-07848	2.0–2.1	0.3 to 0.4	10.1 to 10.5	2, 3, 5	
<i>qOIL-D1a</i>	D1a (1)	Satt580	Satt580	BARC-054393-12560	2.1–3.2	0.3 to 0.4	10.4 to 15.8	3–5, 0	14, 19
<i>qOIL-D1b</i>	D1b (2)	Satt274	Satt274	Staga002	2.1–2.7	–0.3 to –0.4	10.1 to 13.1	1–4, 0	VAL; 13, 14, 17
<i>qOIL-D2</i>	D2 (17)	BARCSOYSSR_17_0690	Sat_222	BARCSOYSSR_17_0690	2.6–5.1	0.2 to 0.4	5.6 to 11.9	3–5, 0	VAL; 14
<i>qOIL-E</i>	E (15)	BARCSOYSSR_15_1073	Satt452	Satt045	2.6–6.4	0.3 to 0.6	9.0 to 26.5	1–5, 0	7
<i>qOIL-G</i>	G (18)	Satt472	Satt472	BARCSOYSSR_18_1699	2.7–5.9	–0.3 to –0.4	5.6 to 13.5	2–5, 0	VAL
<i>qOIL-I-1</i>	I (20)	Satt451	Satt562	Satt614	3.0–3.9	0.3 to 0.5	14.5 to 20.1	2–5, 0	3, 5, 11, 16, 19
<i>qOIL-I-2</i>	I (20)	BARC-057033-14543	BARC-021887-04232	BARC-057049-14550	3.0–4.0	–0.3 to –0.5	15.1 to 20.0	2–5, 0	
<i>qOIL-J</i>	J (16)	BARCSOYSSR_16_1244	Satt431	Sat_350	2.2–3.6	0.3 to 0.5	11.3 to 17.7	1–5, 0	VAL
<i>qOIL-K</i>	K (9)	BARCSOYSSR_09_1171	BARC-014813-01678	BARCSOYSSR_09_1171	3.5–4.8	0.3 to 0.4	7.9 to 11.6	1, 2	VAL
<i>qOIL-N-1</i>	N (3)	Satt009	Satt009	Satt641	2.1–3.3	–0.3 to –0.5	11.6 to 15.9	2, 3, 0	VAL; 14
<i>qOIL-N-2</i>	N (3)	Sat_084	Satt125	Satt584	2.3–3.9	–0.3 to –0.5	12.0 to 18.9	1–5, 0	
<i>qOIL-O</i>	O (10)	Satt581	BARCSOYSSR-10_1394	BARC-050013-09288	3.0–5.3	–0.4 to –0.5	14.5 to 22.9	2–5, 0	VAL
Protein									
<i>qPRO-B1</i>	B1 (11)	BARC-030863-06956	Satt251	BARC-032333-08951	2.1–2.9	0.5 to 0.6	11.0 to 15.7	3–5, 0	VAL; 2
<i>qPRO-C1-1</i>	C1 (4)	BARCSOYSSR_04_1400	Satt338	BARCSOYSSR_04_1400	2.4–3.0	–0.5 to –0.6	11.9 to 14.6	3–5, 0	9
<i>qPRO-C1-2</i>	C1 (4)	BARC-040777-07848	BARC-044521-08714	BARC-064861-18829	2.1–3.3	–0.5 to –0.6	11.2 to 17.1	2, 5, 0	12
<i>qPRO-D1b</i>	D1b (2)	Satt274	Satt274	Sat_289	2.5–3.4	0.5 to 0.7	12.3 to 16.4	1, 2, 0	VAL; 7, 14

Table 2 continued

QTL ^a	LG (Chr #)	Peak marker ^b	Interval		LOD	A.E ^d	R ² (%)	Env ^e	References ^f
			Int-L ^c	Int-R ^c					
<i>qPRO-D2-1</i>	D2 (17)	BARCOYSSR_17_0900	BARCOYSSR_17_0900	Sat_300	2.4–3.4	–0.5 to –0.6	11.7 to 16.3	1–5, 0	VAL; 20
<i>qPRO-D2-2</i>	D2 (17)	BARCOYSSR-17_0621	BARCOYSSR-17_0646	BARCOYSSR-17_0621	2.4–2.7	–0.6 to –0.7	13.2 to 26.0	1–4, 0	VAL
<i>qPRO-I-1</i>	I (20)	BARCOYSSR_20_0079	BARCOYSSR_20_0049	Sat562	4.5–7.9	–0.7 to –1	21.9 to 37.0	1–5, 0	8, 17
<i>qPRO-I-2</i>	I (20)	BARC-041129-07912	Sat587	Sat354	3.3–9.5	–0.6 to –1.1	8.7 to 33.8	1–5, 0	3, 11, 16, 19
<i>qPRO-I-3</i>	I (20)	BARC-057033-14543	BARC-021887-04232	BARC-057049-14550	2.7–6.5	0.6 to 0.7	14.0 to 31.9	1–5, 0	
<i>qPRO-K</i>	K (9)	Sat273	BARCOYSSR_09_1123	BARCOYSSR_09_1171	2.7–3.0	0.5 to 0.6	14.4 to 14.6	3, 4	VAL; 10
<i>qPRO-N</i>	N (3)	Sat584	Sat_084	Sat584	2.2–3.8	0.4 to 0.6	11.1 to 17.9	2–5, 0	10

^a QTL for the same trait detected in different environments with the same or overlapping marker interval was designated as one QTL. The QTLs shown in bold are new ones that were different from those previously reported but validated in the validation population (see “Discussion” section for detail)

^b Peak marker at which the maximum LOD was exhibited

^c Int-L and Int-R indicate the left and right side marker of the interval, respectively

^d A.E additive effect. The positive value indicates that the A02-381100 homozygous alleles increase the values of traits, or the SD02-4-59 alleles decrease the values of traits, and vice versa. The unit of additive effect is kg ha^{–1} for seed yield and percent (%) based on dry seed weight for oil and protein contents

^e Env environments in which QTLs were detected. 1–5 and 0 indicate environment E09AU, E09BF, E10AU, E11VG and E11BF, and the average data over five environments, respectively

^f Indicating that the QTL was confirmed in the validation population of SD02-4-59 × A02-381100 in this study (VAL) and/or has been reported previously in the reference(s): 1. Brummer et al. (1997); 2. Chapman et al. (2003); 3. Chung et al. (2003); 4. Concibido et al. (2003); 5. Csanadi et al. (2001); 6. Guzman et al. (2007); 7. Hyten et al. (2004); 8. Jun et al. (2008); 9. Kabelka et al. (2004); 10. Lee et al. (1996); 11. Nichols et al. (2006); 12. Orf et al. (1999); 13. Panthee et al. (2005); 14. Qi et al. (2011); 15. Reinprecht et al. (2006); 16. Sebolt et al. (2000); 17. Shi et al. (2010); 18. Smalley et al. (2004); 19. Specht et al. (2001); 20. Wang et al. (2004); 21. Yuan et al. (2002). See references part for detailed information

Table 3 Cumulative effect of the two QTLs which were located on the same linkage group and had the same source of favorable alleles for seed yield, oil and protein contents in the mapping population of SD02-4-59 × A02-381100 over five environments

QTL combination	Allele ^a	Mean ^b		QTL combination	Allele ^a	Mean ^b	
<i>qYIE-D2-1</i> + <i>qYIE-D2-2</i>	A + A	2,445.2 ± 754.8	c	<i>qPRO-C1-1</i> + <i>qPRO-C1-2</i>	A + A	41.5 ± 1.4	a
	A + S	2,275.1 ± 734.8	b		A + S	42.3 ± 1.7	b
	S + A	2,338.4 ± 757.0	b		S + A	42.3 ± 1.7	b
	S + S	2,075.2 ± 696.4	a		S + S	43.7 ± 1.8	c
<i>qYIE-I-1</i> + <i>qYIE-I-2</i>	A + A	2,410.3 ± 785.0	c	<i>qPRO-D2-1</i> + <i>qPRO-D2-2</i>	A + A	41.8 ± 1.6	a
	A + S	2,226.0 ± 740.7	ab		A + S	43.3 ± 1.4	c
	S + A	2,333.1 ± 768.6	bc		S + A	42.7 ± 1.8	b
	S + S	2,117.8 ± 653.5	a		S + S	43.2 ± 1.8	c
<i>qOIL-N-1</i> + <i>qOIL-N-2</i>	A + A	17.9 ± 1.2	a	<i>qPRO-I-1</i> + <i>qPRO-I-2</i>	A + A	41.6 ± 1.5	a
	A + S	18.6 ± 1.2	b		A + S	42.6 ± 1.5	b
	S + A	18.4 ± 1.0	abc		S + A	42.0 ± 1.4	b
	S + S	18.9 ± 1.1	c		S + S	43.6 ± 1.7	c

^a A Homozygous alleles for A02-381,100, and S Homozygous alleles for SD02-4-59, respectively

^b Means with different letters within the same two-QTL combination were significantly different at $P < 0.05$ by LSD

detected in three environments and the average data over all environments, explaining 14.4–20.2 %, 10.9–24.0 % and 15.9–37.3 % of the phenotypic variation, respectively. The QTL on linkage group O (*qYIE-O*) was detected in three environments, but was not detected by the combined data. The QTLs on linkage group F, I and K (*qYIE-F*, *qYIE-I-1* and *qYIE-K*) were detected in two environments and the average data over all environments, accounting for 11.1–18.2 %, 19.0–30.6 % and 13.4–14.5 % of the total variation, respectively. The QTLs on linkage group D2, E, J and M (*qYIE-D2-1*, *qYIE-E*, *qYIE-J* and *qYIE-M*) were detected in one environment and the average data. The QTL on linkage group G (*qYIE-G*) was detected in two environments, explaining 11.3 % and 13.2 % of the total variation.

Sixteen QTLs for oil content were identified on linkage groups A1, B1, C1, D1a, D1b, D2, E, G, I, J, K, N and O by CIM and/or IM analysis (Table 2). Two QTLs (*qOIL-N-1* and *qOIL-N-2*) with the same source of the favorable allele were mapped on linkage group N. The results of ANOVA for the comparison of QTL combinations showed that there was a significant cumulative effect for them (Table 3), indicating that these two QTLs were independent of each other. The QTLs on linkage groups E, J and N (*qOIL-E*, *qOIL-J* and *qOIL-N-2*) were consistently detected in all the five environments and the average data, explaining 9.0–26.5 %, 11.3–17.7 % and 12.0–18.9 % of the total variation, respectively. The QTLs on linkage groups D1b, G, I and O (*qOIL-D1b*, *qOIL-G*, *qOIL-I-1* and *qOIL-I-2*, and *qOIL-O*) were consistently detected in four environments and the average data. The QTLs on linkage groups B1, C1, D1a and D2 (*qOIL-B1*, *qOIL-C1-1*, *qOIL-D1a*, and *qOIL-D2*) were detected in three environments and the average data. The QTLs on linkage groups C1 and K (*qOIL-C1-2*

and *qOIL-K*) were identified in three and two environments, respectively, but were not mapped in the average data. The QTLs on linkage groups A1 and N (*qOIL-A1* and *qOIL-N-1*) were detected in two environments and the average data, accounting for 6.6–23.0 % and 11.6–15.9 % of the variation, respectively.

Eleven QTLs for protein content were identified on linkage groups B1, C1, D1b, D2, I, K and N by CIM and/or IM analysis (Table 2). Two QTLs with the same source of favorable alleles were mapped on each of the linkage groups C1, D2 and I. The results of group comparisons for QTL combinations indicated that the two QTLs on the linkage group C1, D2 or I exhibited significant cumulative effects (Table 3), and thus, in each case, they were independent of each other. One QTL on linkage group D2 (*qPRO-D2-1*) and three QTLs on linkage group I (*qPRO-I-1*, *qPRO-I-2* and *qPRO-I-3*) were consistently detected in all the five environments and the average data, explaining 11.7–16.3 %, 21.9–37.0 %, 8.7–33.8 % and 14.0–31.9 % of the phenotypic variation, respectively. The QTLs on linkage groups D2 and N (*qPRO-D2-2* and *qPRO-N*) were consistently identified in four environments and the average data, explaining 13.2–26.0 % and 11.1–17.9 % of the phenotypic variation, respectively. The QTLs on linkage groups B1 and C1 (*qPRO-B1* and *qPRO-C1-1*) were repeatedly detected in three individual environments and the average data. Another QTL on linkage group C1 (*qPRO-C1-2*) and the QTL on linkage group D1b (*qPRO-D1b*) were detected in two environments and the average data, accounting for 11.2–17.1 % and 12.3–16.4 % of the phenotypic variation, respectively. The QTL on linkage group K (*qPRO-K*) was detected in two individual environments, explaining 14.4–14.6 % of the variation. This QTL was the

only one for protein that could not be detected in the average data in this study.

Validation of the QTLs in the RIL population of SD02-911 × SD00-1501

Of 1,077 SSR markers screened, 249 SSR markers showed polymorphism between the two parents of the validation population. Two hundred and six polymorphic SSR markers were located on the 16 linkage groups that carried QTLs associated with seed yield, oil and/or protein content in the mapping population (Table 2). Thus the 206 polymorphic SSR markers were selected to genotype all the RILs in the validation population. SMA and ANOVA for comparisons between different allele groups over all environments indicated that 18 QTLs for seed yield, oil and protein contents were confirmed in the validation population (Table 4).

Five QTLs for seed yield on the linkage groups B2, D2, F, M and O were confirmed in the validation population in the environment E12BK and the average data over three environments (Table 4). The favorable alleles for all the QTLs in the validation population were derived from the parent SD02-911, except for the QTL on linkage group F, which was inherited from the parent SD00-1501. The QTL on linkage group O showed a larger additive effect in the validation population than in the mapping population, while other validated QTLs showed similar magnitudes of additive effects in both populations.

Eight QTLs for oil content on linkage groups C1, D1b, D2, G, J, K, N and O were confirmed in the validation population (Table 4). The QTLs on linkage groups D2, G and N (*qOIL-D2*, *qOIL-G* and *qOIL-N*) were confirmed in all three environments and the combined data. The QTL on linkage group K (*qOIL-K*) was confirmed in two environments and the combined data in the validation population. The QTLs on linkage groups C1, D1b, J and O (*qOIL-C1-1*, *qOIL-D1b*, *qOIL-J* and *qOIL-O*) were confirmed in the environment E12BK and the average data. The favorable alleles for six QTLs detected in the validation population were derived from the parent SD02-911, while the QTLs on linkage groups D1b and O (*qOIL-D1b* and *qOIL-O*) were derived from the parent SD00-1501.

Five QTLs for protein content on linkage groups B1, D1b, D2 and K were confirmed in the validation population (Table 4). Two QTLs on linkage group D2 (*qPRO-D2-1* and *qPRO-D2-2*) and the QTLs on linkage group B1 and D1b (*qPRO-B1* and *qPRO-D1b*) were detected in two environments and the average data in the validation population. The QTL on linkage group K (*qPRO-K*) was confirmed in one environment and the average data. The favorable alleles for all the protein QTLs confirmed in the validation population were derived from the same source, i.e. the parent SD00-1501.

Discussion

Heritability of seed yield, oil and protein contents

Seed yield is the most important trait, and oil and protein contents are the economically important quality traits in soybean. There were significant differences in all three traits among the RILs in the two populations in this study ($P < 0.01$, Table 1). Previous studies indicated that the estimates of heritability for oil and protein contents varied from 0.07 to 0.89 and 0.56 to 0.92, respectively, depending on the populations and the environments (Lee et al. 1996; Brummer et al. 1997; Csanadi et al. 2001; Chung et al. 2003; Hyten et al. 2004). In our study, the estimated heritability was 0.94 for both oil and protein content in the mapping population, and 0.84 and 0.71 in the validation population, respectively (Table 1), which was comparable to the results of most previous studies. Yuan et al. (2002) reported a heritability of 0.47 for yield across four environments with a range from 0.25 to 0.50 within specific environments. Guzman et al. (2007) reported a range of 0.77–0.87 for the heritability of yield across multiple environments, depending on the population. Palomeque et al. (2009a) estimated the heritability of yield in soybean based on multiple-environment data and found that the estimates were 0.64 in China and 0.89 in Canada, respectively. In our study, the heritability across environments for yield was 0.65 in the mapping population and 0.19 in the validation population (Table 1), which was comparable to those reported by Yuan et al. (2002) and that in China by Palomeque et al. (2009a), but lower than those reported by Guzman et al. (2007) and that in Canada by Palomeque et al. (2009a). Compared with the validation population, the mapping population showed higher estimates of heritability for seed yield, oil and protein contents. This should be attributed to the differences in population size and experimental error between the two populations.

QTLs for yield

In this study, 12 QTLs for seed yield were detected in the mapping population and 5 of them were verified in the validation population (Table 2). Orf et al. (1999) and Smalley et al. (2004) reported a yield QTL linked to the marker Satt066 on linkage group B2. Concibido et al. (2003) identified a yield-enhancing QTL in the same region on linkage group B2 from *Glycine soja* (Siebold and Zucc.) in a population derived from the cross of HS-1 × PI 407305. Guzman et al. (2007) also mapped this yield QTL to the same region on linkage group B2. Our results further confirmed the previously reported QTL on linkage group B2, because the yield QTL on linkage group B2 detected in this study shared the same marker (Satt534) which was associated

Table 4 Means of RILs carrying different alleles of the validated QTLs for seed yield (kg ha^{-1}), oil and protein contents (%) in the mapping population of SD02-4-59 $\times$ A02-381100 and in the validation population of SD02-911 $\times$ SD00-1501

QTL	Mapping population		Validation population						
	A ^a	S ^a	Difference	Validated marker	LOD ^b	Env. ^c	B ^d	D ^d	Difference
Yield									
<i>qYIE-B2</i>	2,382.7 ± 791.8	2,204.5 ± 699.4	178.2***	Satt534	6.4	3, 0	2,574.1 ± 364.6	2,431.4 ± 394.9	142.7***
<i>qYIE-D2-1</i>	2,366.4 ± 748.6	2,212.5 ± 738.9	153.9***	Sct_192	2.6	3, 0	2,541.0 ± 372.8	2,414.2 ± 409.9	126.8***
<i>qYIE-F</i>	2,139.3 ± 665.7	2,357.9 ± 774.8	-218.5***	Sat_298	3.8	3, 0	2,315.5 ± 413.8	2,536.3 ± 374.7	-220.8***
<i>qYIE-M</i>	2,389.3 ± 750.9	2,209.1 ± 732.0	180.2***	Sat_226	5.4	3, 0	2,545.5 ± 373.7	2,360.3 ± 385.9	185.2***
<i>qYIE-O</i>	2,323.1 ± 793.5	2,208.6 ± 656.3	114.5**	Satt243	4.3	3, 0	2,546.3 ± 373.1	2,375.6 ± 399.8	170.7***
Oil									
<i>qOIL-C1-1</i>	18.1 ± 1.2	18.7 ± 1.2	-0.6***	Sat_140	3.7	3, 0	21.1 ± 0.8	20.6 ± 0.9	0.5***
<i>qOIL-D1b</i>	18.2 ± 1.2	18.8 ± 1.2	-0.6***	Satt459	2.6	3, 0	20.8 ± 0.6	21.1 ± 0.9	-0.3*
<i>qOIL-D2</i>	18.7 ± 1.2	18.1 ± 1.2	0.6***	Sat_222	27.7	1-3, 0	21.4 ± 0.7	20.4 ± 0.7	1.0***
<i>qOIL-G</i>	18.1 ± 1.2	18.8 ± 1.1	-0.7***	BARCOYSSR_18_1758	34.9	1-3, 0	21.4 ± 0.7	20.4 ± 0.7	1.0***
<i>qOIL-J</i>	18.8 ± 1.1	18.0 ± 1.3	0.7***	Sat_350	4.0	3, 0	21.1 ± 0.8	20.6 ± 0.9	0.5***
<i>qOIL-K</i>	18.7 ± 1.3	18.3 ± 1.2	0.4***	BARCOYSSR_09_1171	4.4	2,3, 0	21.4 ± 0.7	20.9 ± 0.9	0.5***
<i>qOIL-N-1</i>	18.2 ± 1.2	18.9 ± 1.1	-0.7***	Satt009	18.3	1-3, 0	21.3 ± 0.8	20.5 ± 0.6	0.8***
<i>qOIL-O</i>	18.1 ± 1.2	18.9 ± 1.1	-0.8***	Satt581	2.8	3, 0	20.5 ± 1.0	21.1 ± 0.8	-0.6***
Protein									
<i>qPRO-B1</i>	43.0 ± 1.8	41.8 ± 1.7	1.2***	Satt251	5.4	2,3, 0	39.3 ± 1.0	39.9 ± 1.3	-0.6***
<i>qPRO-D1b</i>	42.7 ± 1.9	41.8 ± 1.5	0.9***	Staga002	20.7	1,3, 0	39.5 ± 1.0	41.1 ± 1.6	-1.7***
<i>qPRO-D2-1</i>	41.9 ± 1.6	42.9 ± 1.9	-1.0***	Sat_300	16.2	2,3, 0	39.5 ± 1.0	41.0 ± 1.7	-1.6***
<i>qPRO-D2-2</i>	42.1 ± 1.7	43.2 ± 1.8	-1.1***	BARCOYSSR_17_0621	13.4	2,3, 0	39.4 ± 1.0	40.4 ± 1.5	-1.0***
<i>qPRO-K</i>	42.7 ± 2.0	42.0 ± 1.5	0.7***	BARCOYSSR_09_1171	2.7	3, 0	39.2 ± 0.8	39.8 ± 1.3	-0.6***

^a A and S indicate the homozygous allelic means of A02-381100 and SD02-4-59 over five environments, respectively^b LOD score from the results of single marker analysis of seed yield, oil and protein contents over three environments in the validation population of SD02-911 $\times$ SD00-1501^c Env environments in which the associated SSR markers were confirmed in the validation population of SD02-911 $\times$ SD00-1501. 1-3 and 0 indicate environment E11AU, E12VG and E12BK, and the average data over three environments, respectively^d B and D indicate the homozygous allelic means of SD02-911 and SD00-1501 over three environments, respectively*, **, and *** Significant at $P < 0.05$, 0.01 and 0.001, respectively, based on ANOVA for two groups of RILs carrying different marker alleles over all three environments

with the QTL reported by Concibido et al. (2003). In this study, two QTLs for seed yield (*qYIE-D2-1* and *qYIE-D2-2*) were detected on linkage group D2 with peak marker Sct_192 and BARCSOYSSR_17_0900, respectively. Smalley et al. (2004) reported that SSR markers Sctt008, Satt135 and Satt311 were significantly associated with seed yield. Referring to the public genetic map (Song et al. 2010, supplementary Table 1), Sct_192 is at 13.3 cM, Sctt008 is at 4.5 cM and Satt135 is at 25.5 cM. Thus, *qYIE-D2-1* is about 10 cM away from markers Sctt008 and Satt135. Additional studies will help to determine if *qYIE-D2-1* is the same QTL as previously reported or not. On the integrated soybean genetic map version 2010 (Song et al. 2010), the marker BARCSOYSSR_17_0900 is located close to Satt311 (<2 cM), indicating that *qYIE-D2-2* confirmed the QTL linked to Satt311 (Smalley et al. 2004).

Smalley et al. (2004) identified three markers (Satt127, Satt239 and Satt270) on linkage group I were associated with seed yield. The peak marker for *qYIE-I-2*, BARCSOYSSR_20_0750, is located close to Satt270 (<2 cM) on the integrated linkage map (Song et al. 2010), indicating that *qYIE-I-2* should be the same QTL as the one reported by Smalley et al. (2004). The peak marker for *qYIE-I-1* is located far away (>15 cM) from the QTLs on linkage group I reported by Yuan et al. (2002) and Smalley et al. (2004), suggesting that *qYIE-I-1* might be a new QTL for seed yield. The QTL on linkage group J (*qYIE-J*) detected in this study is consistent with the QTL reported by Guzman et al. (2007), because both QTL regions covered the same marker Satt547. Previous studies have repeatedly identified a QTL in the 35.0–45.0 cM region of linkage group K on the integrated linkage map (Specht et al. 2001; Yuan et al. 2002; Smalley et al. 2004; Kabelka et al. 2004; Wang et al. 2004; Guzman et al. 2007). It was further confirmed in our study since the peak mark (Satt273) for *qYIE-K* was also located within that region on the integrated linkage map (Song et al. 2010).

The QTL (*qYIE-M*) on linkage group M identified in this study shared the same marker Satt540 with the QTL identified by Smalley et al. (2004), which was also detected by Orf et al. (1999), Specht et al. (2001), Kabelka et al. (2004) and Wang et al. (2004). Smalley et al. (2004) found a QTL associated with Satt331 on linkage group O. Referring to the consensus linkage map (Song et al. 2004), the peak marker for the QTL on linkage group O (*qYIE-O*) identified in this study is about 2.6 cM away from Satt331. Therefore, the QTL *qYIE-O* should be the same as reported by Smalley et al. (2004). QTLs on linkage groups E, F and G have been previously reported (Orf et al. 1999; Reyna and Sneller 2001; Specht et al. 2001; Kabelka et al. 2004; Wang et al. 2004). However, the QTLs *qYIE-E*, *qYIE-F* and *qYIE-G* identified in this study were located far away (i.e. >15 cM) from the QTLs identified in other studies

according to the integrated linkage map (Song et al. 2010). In summary, therefore, *qYIE-E*, *qYIE-F* and *qYIE-G* as well as *qYIE-D2-1* and *qYIE-I-1* discussed above might be new QTLs for seed yield.

QTLs for oil content

For oil concentration, 16 QTLs were detected in the mapping population and 8 of them were verified in the validation population (Table 2). Brummer et al. (1997) identified an oil QTL linked to a restriction fragment length polymorphism (RFLP) marker A975-1 on linkage group A1. Orf et al. (1999) and Specht et al. (2001) confirmed this QTL associated with SSR markers Satt258, Satt225 and Satt174 in a region of 77.1–82.8 cM on linkage group A1 on the integrated linkage map (Song et al. 2010). In our study, the marker interval for *qOIL-A1* overlapped the region of 77.1–82.8 cM on linkage group A1, suggesting that *qOIL-A1* confirmed the QTL reported previously (Brummer et al. 1997; Orf et al. 1999; Specht et al. 2001). Qi et al. (2011) identified a QTL for oil content on linkage group B1 with the marker interval Satt197–Satt251, which overlapped the marker interval for *qOIL-B1* identified in this study. Thus, *qOIL-B1* should be the same QTL on linkage group B1 as reported by Qi et al. (2011).

Specht et al. (2001) mapped a QTL for oil content associated with the marker Satt468 on linkage group D1a, and Qi et al. (2011) confirmed this QTL in the population of Charleston × Dong nong594. The marker interval for the QTL *qOIL-D1a* identified in this study covered the marker Satt468 on the integrated linkage map (Song et al. 2010), indicating that *qOIL-D1a* should be the same QTL reported previously (Specht et al. 2001; Qi et al. 2011). Panthee et al. (2005), Shi et al. (2010) and Qi et al. (2011) mapped a QTL for oil content in a similar region with marker Satt274 and Satt459 on linkage group D1b. On the integrated linkage map (Song et al. 2010), the marker interval of *qOIL-D1b* identified in this study covered Satt274 and Satt459. Therefore, our results further confirmed the known QTL on linkage group D1b (Panthee et al. 2005; Shi et al. 2010; Qi et al. 2011). On linkage group D2, a QTL for oil content was identified by Qi et al. (2011). Its marker interval Sat_001–Sat_114 on the integrated linkage map (Song et al. 2010) is similar to the marker interval of *qOIL-D2* identified in this study. Thus, our results confirmed the QTL on linkage D2 reported by Qi et al. (2011). Hyten et al. (2004) identified a QTL for oil content associated with marker Satt268 on linkage group E. The peak marker for *qOIL-E*, BARCSOYSSR_15_1073, is located close to Satt268 (<2 cM) on the integrated linkage map, which confirmed the QTL reported by Hyten et al. (2004).

Previous studies have repeatedly confirmed a QTL for oil content in the region of 20–30 cM on linkage group I

(Sebolt et al. 2000; Specht et al. 2001; Csanadi et al. 2001; Chung et al. 2003; Nichols et al. 2006). Two QTLs for oil content were detected on linkage group I in this study. The marker interval for *qOIL-I-1* was also located within the region of 20–30 cM on linkage group I, indicating *qOIL-I-1* should be the same QTL reported previously (Sebolt et al. 2000; Specht et al. 2001; Csanadi et al. 2001; Chung et al. 2003; Nichols et al. 2006). Since no common markers were used, a comparison with previous studies could not be made, thus we could not determine if *qOIL-I-2* is a new QTL for oil content. Qi et al. (2011) identified a QTL for oil content on linkage group N with the marker interval Satt009–Satt530. On the integrated linkage map (Song et al. 2010), this interval shares the same interval marker Satt009 for one of the two QTLs (*qOIL-N-1*) detected in this study. However, for another QTL on linkage group N (*qOIL-N-2*), no marker close to its interval was identified in previous studies. Therefore, *qOIL-N-1* confirmed the previous report (Qi et al. 2011), while *qOIL-N-2* should be a new QTL for oil content. In addition, QTLs for oil content have been identified on linkage groups C1 (Orf et al. 1999; Kabelka et al. 2004; Fasoula et al. 2004), G (Brummer et al. 1997; Specht et al. 2001; Qi et al. 2011), J (Chr 16) (Specht et al. 2001; Kabelka et al. 2004), K (Brummer et al. 1997; Csanadi et al. 2001), and O (Panthee et al. 2005; Shi et al. 2010). Referring to the integrated linkage map (Song et al. 2010), however, the marker intervals of these QTLs are located far away (i.e. >15 cM) from the intervals of the QTLs identified in this study. Therefore, the QTLs on linkage groups C1 (*qOIL-C1-1* and *qOIL-C1-2*), G (*qOIL-G*), I (*qOIL-I-2*), J (*qOIL-J*), K (*qOIL-K*), N (*qOIL-N-2*) and O (*qOIL-O*) identified in this study appear to be new QTLs for oil content.

QTLs for protein content

In this study, 11 QTLs for protein content were detected in the mapping population and 5 of them were verified in the validation population (Table 2). On linkage group B1, Chapman et al. (2003) identified a QTL for protein content associated with marker Satt251 in an F₂ and F_{4,6} soybean population of Essex × Williams. The QTL *qPRO-B1* shared the same marker Satt251 and thus confirmed the QTL reported by Chapman et al. (2003). Kabelka et al. (2004) and Orf et al. (1999) detected two QTLs for protein content on linkage group C1 associated with the markers Satt338 and Satt578, respectively. Referring to the integrated linkage map (Song et al. 2010), the marker intervals for *qPRO-C1-1* and *qPRO-C1-2* contained Satt338 and Satt578, respectively. Therefore, these two QTLs validated the QTLs reported previously (Kabelka et al. 2004; Orf et al. 1999). On linkage group D1b, a QTL for protein content associated with the SSR marker Satt459 was identified

by Hyten et al. (2004) and confirmed by Qi et al. (2011). Our results (*qPRO-D1b*) further confirmed this QTL since the marker Satt459 is located within the marker interval of *qPRO-D1b*. Reinprecht et al. (2006) reported a QTL for protein content on linkage group D2 with marker Satt389. On the integrated linkage map (Song et al. 2010), the marker Satt389 is located at 68.2 cM, which is very close to the marker interval or region (71–76 cM) for *qPRO-D2-1*. Therefore, *qPRO-D2-1* identified in this study might be the same as the QTL reported by Reinprecht et al. (2006). No marker being located close to the marker interval of *qPRO-D2-2* has been reported previously to be significantly associated with protein content, suggesting that *qPRO-D2-2* should be a new QTL for protein content.

Jun et al. (2008) and Shi et al. (2010) reported a QTL for protein content on linkage group I associated with the marker Satt571, which is located at 14.9 cM on the integrated linkage map (Song et al. 2010). The interval for *qPRO-I-1* covered the marker Satt571, and thus we suppose that *qPRO-I-1* is the same QTL (Jun et al. 2008; Shi et al. 2010). Another QTL for protein content on linkage group I, which is located in the region of 29.6–31.5 cM on the integrated linkage map, was repeatedly verified by different research groups and proven to be the most stable QTL for protein content (Sebolt et al. 2000; Specht et al. 2001; Chung et al. 2003; Nichols et al. 2006). Our results further validated this QTL, because the marker interval for *qPRO-I-2* identified in this study is at 31.5–46.2 cM on the integrated linkage map. Since no comparable markers were used, it could not be assured that *qPRO-I-3* identified in this study is a new QTL for protein content. Additional studies will be needed to verify this conclusion. Lee et al. (1996) identified two protein QTLs associated with marker A065_3 or gac34-2, which were located on linkage groups K and N, respectively. Referring to the GmComposite2003 map (Soybase), the marker intervals for *qPRO-K* and *qPRO-N* detected in this study covered the marker A065_3 and gac34-2, respectively. It indicated that our results confirmed the reported QTLs on linkage groups K and N (Lee et al. 1996). As discussed above, however, two QTLs (*qPRO-D2-2* and *qPRO-I-3*) identified in this study were not reported previously. Therefore, we would suppose that these two QTLs are new ones for protein content in soybean.

Relationship between seed yield, oil and protein content

Previous studies indicated that there were negative correlations between seed yield or oil content and protein content, but positive correlation between yield and oil (Burton 1987; Lee et al. 1996; Brummer et al. 1997; Chung et al. 2003; Panthee et al. 2005; Yesudas et al. 2013). Phenotypic analysis of correlation in both mapping and validation population

Fig. 1 Map locations of QTLs overlapped or closely located for seed yield, oil and protein contents in the mapping population of SD02-4-59 × A02-381100

also exhibited similar results (data not shown). Based on the QTL information as described above, it is noted that some QTLs associated with one trait were also detected for other one or two traits in the mapping population (Fig. 1), suggesting that these QTLs might have pleiotropic effects or might be closely linked. The QTLs on linkage group D2, I and K for yield, oil and protein content, the QTLs on linkage group B1, C1, D1b and N for oil and protein content, and the QTLs on linkage group E and O for oil content and yield were situated in close proximity to one another (the marker intervals were overlapped or located less than

2 cM away) (Table 2; Fig. 1). Of the closely located or overlapped QTLs, the QTLs on linkage group D2 (*qYIE-D2-2*, *qOIL-D2* and *qPRO-D2-1*) and I (*qYIE-I-1*, *qYIE-I-2*, *qOIL-I*, *qPRO-I-1* and *qPRO-I-2*) exhibited positive effects on yield and oil, but negative effects on protein. The QTLs on linkage group B1 (*qOIL-B1* and *qPRO-B1*), C1 (*qOIL-C1-2* and *qPRO-C1-2*), D1b (*qOIL-D1b* and *qPRO-D1b*) and N (*qOIL-N-2* and *qPRO-N*) exhibited opposite genetic effects on oil and protein. The QTLs on linkage group E (*qYIE-E* and *qOIL-E*) for yield and oil content exhibited positive effects on both traits. The effects of these QTLs

showed high consistency with the results of phenotypic correlation analysis. Contrarily, however, the effects of QTLs on linkage group O (*qYIE-O* and *qOIL-O*) were positive for yield but negative for oil content, and the QTLs on linkage group K (*qOIL-K* and *qPRO-K*) exhibited positive effects on both oil and protein contents. In addition, there were also several QTLs which were specific for yield, oil or protein (Table 2), suggesting that there might be possibility to improve the traits simultaneously to some extent by pyramid the favorable alleles. Molecular information may provide a better understanding of the relationships among traits, like yield, oil and protein contents in soybean.

Population size and QTL identification

In general, large populations are more effective in detecting QTLs, especially for minor-effect QTLs. Along with an increase of population sizes, however, control of the experimental error in phenotyping might become less easy, and the consistency between different environments could decrease with increased population sizes. Using the same field experimental design, a large population size may lower the level of QTL consistency for lower heritability traits. Therefore, appropriate population sizes depend on different factors and considerations. In practical studies, the sample sizes varied considerably, from 60 to 380 (Melchinger et al. 2000). A small population with appropriate variation should be considered effective if the results can be repeatedly verified in multiple environments and/or across multiple populations (Cornelius et al. 2005; Yesudas et al. 2013). As discussed previously (Wang et al. 2012, 2014), the mapping population in this study was small, consisting of 87 F₅-derived RILs, and might be less effective in detecting minor-effect QTLs than a larger population. However, phenotyping for the three traits of interest (seed yield, oil and protein contents) was conducted in five independent environments, and the QTLs identified were repeatedly detected in multiple environments and in the combined data over all environments. Moreover, to verify the QTLs identified in the mapping population, a large validation population consisting of 196 RILs was phenotyped in three independent environments and genotyped with all polymorphic SSR markers which were also associated with the QTLs identified in the mapping population. Consequently, about half of the QTLs identified in the mapping population were confirmed in the validation population both by SMA in at least one environment and in the average data over all three environments and by ANOVA over all environments. Of these validated QTLs, eight QTLs were not reported previously, and thus they should be new ones as highlighted in Table 2. In addition, as discussed above, of the 39 QTLs identified in this study, 24 QTLs (seven for yield, eight for oil content and nine for protein content) have been detected

in previous studies (Table 2). This means that most of the QTLs identified in our study were consistent with the results of other studies. Therefore, the results of this study are reliable and informative, though the mapping population was relatively small. However, the proportions of total variation explained by single QTLs were close to or greater than 10 % in most cases. Thus, the genetic effects of the QTLs identified in the mapping population might be overestimated due to the small population size (Beavis 1994).

The effectiveness of QTL identification and validation is also dependent on the heritability of traits of interest. In this study, the QTLs for seed yield were less repeatedly or less consistently detected in multiple environments than those for oil and/or protein. It is understandable because the estimates of heritability for yield were obviously smaller than those for oil and protein (Table 1). Compared with the mapping population, the validation population exhibited a lower level of QTL consistency or repeatability because of larger population size and lower heritability. Controlling the experimental errors helps to enhance repeatability and is important especially for QTL mapping with a large population.

In conclusion, 12 QTLs for seed yield, 16 QTLs for oil content and 11 QTLs for protein content were consistently detected in multiple environments and/or the average data over all environments in the mapping population. Of the QTLs detected in the mapping population, five QTLs for seed yield, eight QTLs for oil content and five QTLs for protein content were confirmed in the validation population by both SMA and ANOVA over all environments. Eight of these validated QTLs were newly identified and are first reported here. Furthermore, seven QTLs for seed yield, eight QTLs for oil content and nine QTLs for protein content also verified the previously reported QTLs. Therefore, most of the QTLs identified in the mapping population were either confirmed in different studies or validated in a different population, and will be useful for breeding higher yield and better quality soybean cultivars. In addition, some QTLs were specific for seed yield, oil content or protein content, while some QTLs associated with one trait were also overlapped with or closely linked to the QTLs of other one or two traits. All this information provides a better understanding of the relationships among seed yield, oil and protein contents, and will help effectively and efficiently improve yield potential and nutritional quality in soybean.

Acknowledgments This research was supported in part by the USDA-NIFA/CSREES Hatch project, the USB Seed Composition Research project via USDA/ARS, and the South Dakota Soybean Research & Promotion Council project. Genotyping of three *GmFAD3* genes was performed by Dr. Kristin Bilyeu's lab, USDA/ARS in Columbia, MO. Jiaoping Zhang, Siddhi Bhusal and Nicholas Hall assisted in field experiments and seed composition

analysis. The parental line A02-381100 was provided by Dr. Walter Fehr, Iowa State University.

References

- Bachlava E, Dewey RE, Burton JW, Cardinal AJ (2009) Mapping and comparison of quantitative trait loci for oleic acid seed content in two segregating soybean populations. *Crop Sci* 49:433–442
- Beavis WD (1994) The power and deceit of QTL experiments: lessons from comparative QTL studies. In: 49th annual corn and sorghum industry research conference. American Seed Trade Association, Washington, DC
- Brummer EC, Graef GL, Orf J, Wilcox JR, Shoemaker RC (1997) Mapping QTL for seed protein and oil content in eight soybean populations. *Crop Sci* 37:370–378
- Burton JW (1987) Quantitative genetics: results relevant to soybean breeding. In Wilcox JR (ed) Soybeans: improvement, production and uses. 2nd edn. Agron Monogr 16. ASA, CSSA, and SSSA, Madison
- Chapman A, Pantalone VR, Ustun A, Allen FL, Landau-Ellis D, Triganio RN, Gresshoff PM (2003) Quantitative trait loci for agronomic and seed quality traits in an F_2 and $F_{4:6}$ soybean population. *Euphytica* 129:387–393
- Chiari L, Piovesan ND, Naoe LK, Jose IC, Viana JMS, Moreira MA, De Barros EG (2004) Genetic parameters relating isoflavone and protein content in soybean seeds. *Euphytica* 138:55–60
- Chung J, Babka HL, Graef GL, Staswick PE, Lee DJ, Cregan PB, Shoemaker RC, Specht JE (2003) The seed protein, oil, and yield QTL on soybean linkage group I. *Crop Sci* 43:1053–1067
- Clemente TE, Cahoon EB (2009) Soybean oil: genetic approaches for modification of functionality and total content. *Plant Physiol* 151:1030–1040
- Concibido VC, Vallee BL, McIaird P, Pineda N, Meyer J, Hummel L, Yang J, Wu K, Delannay X (2003) Introgression of a quantitative trait locus for yield from *Glycine soja* into commercial soybean cultivars. *Theor Appl Genet* 106:575–582
- Cornelius B, Chen P, Chen Y, de Leon N, Shannon JG, Wang D (2005) Identification of QTLs underlying water-logging tolerance in soybean. *Mol Breed* 16:103–112
- Csanadi G, Vollmann J, Stift G, Lelley T (2001) Seed quality QTLs identified in a molecular map of early maturing soybean. *Theor Appl Genet* 103:912–919
- Diers BW, Keim P, Fehr WR, Shoemaker RC (1992) RFLP analysis of soybean seed protein and oil content. *Theor Appl Genet* 83:608–612
- Fasoula VA, Harris DK, Boerma HR (2004) Validation and designation of quantitative trait loci for seed protein, seed oil, and seed weight from two soybean populations. *Crop Sci* 44:1218–1225
- Fehr WR (1987) Principles of cultivar development—Volume 1, Theory and Technique. Macmillan Publishing Company, New York
- Guzman PS, Diers BW, Neece DJ, Martin SK St, LeRoy AR, Grau CR, Hughes TJ, Nelson RL (2007) QTL associated with yield in three backcross-derived populations of soybean. *Crop Sci* 47:111–122
- Hyten DL, Pantalone VR, Sams CE, Saxton AM, Landau-Ellis D, Stefaniak TR, Schmidt ME (2004) Seed quality QTL in a prominent soybean population. *Theor Appl Genet* 109:552–561
- Jiang G-L, Shi JR, Ward RW (2007) QTL analysis of resistance to Fusarium head blight in the novel wheat germplasm CJ 9306. I Resistance to fungal spread. *Theor Appl Genet* 116:3–13
- Jun T-H, Van K, Kim MY, Lee S-H, Walker DR (2008) Association analysis using SSR markers to find QTL for seed protein content in soybean. *Euphytica* 162:179–191
- Kabelka EA, Diers BW, Fehr WR, LeRoy AR, Baianu IC, You T, Neece DJ, Nelson RL (2004) Putative alleles for increased yield from soybean plant introductions. *Crop Sci* 44:784–791
- Lee SH, Bailey MA, Mian MAR, Carter TE, Shipe ER, Ashley DA, Parrott WA, Hussey RS, Boerma HR (1996) RFLP loci associated with soybean seed protein and oil content across populations and locations. *Theor Appl Genet* 93:649–657
- Li D, Pfeiffer TW, Cornelius PL (2008) Soybean QTL for yield and yield components associated with *Glycine soja* alleles. *Crop Sci* 48:571–581
- Melchinger AE, Utz HF, Schon CC (2000) From Mendel to Fisher. The power and limits of QTL mapping for quantitative traits. *VortPflanzenzüchtg* 48:132–142
- Nichols DM, Golver KD, Carlson SR, Specht JE, Diers BW (2006) Fine mapping of a seed protein QTL on soybean linkage group I and its correlated effects on agronomic traits. *Crop Sci* 46:834–839
- 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
- Palomeque LC, Jun LL, Li W, Hedges B, Cober ER, Rajcan I (2009a) QTL in mega-environments: I. Universal and specific seed yield QTL detected in a population derived from a cross of high-yielding adapted × high-yielding exotic soybean lines. *Theor Appl Genet* 119:417–427
- Palomeque LC, Jun LL, Li W, Hedges B, Cober ER, Rajcan I (2009b) QTL in mega-environments: II. Agronomic trait QTL co-localized with seed yield QTL detected in a population derived from a cross of high-yielding adapted × high-yielding exotic soybean lines. *Theor Appl Genet* 119:429–436
- Palomeque L, Liu LJ, Li W, Hedges BR, Cober ER, Smid MP, Lukens L, Rajcan I (2010) Validation of mega-environment universal and specific QTL associated with seed yield and agronomic traits in soybeans. *Theor Appl Genet* 120:997–1003
- Panthee DR, Pantalone VR, West DR, Saxton AM, Sams CE (2005) Quantitative trait loci for seed protein and oil concentration, and seed size in soybean. *Crop Sci* 45:2015–2022
- Panthee DR, Pantalone VR, Saxton AM (2006) Modifier QTL for fatty acid composition in soybean oil. *Euphytica* 152:67–73
- Qi Z-M, Wu Q, Han X, Sun Y-N, Du X-Y, Liu C-Y, Jiang H-W, Hu G-H, Chen Q-S (2011) Soybean oil content QTL mapping and integrating with meta-analysis method for mining genes. *Euphytica* 179:499–514
- Reinprecht Y, Poysa VW, Yu K, Rajcan I, Ablett GR, Pauls KP (2006) Seed and agronomic QTL in low linolenic acid, lipoxygenase-free soybean (*Glycine max* (L.)Merrill) germplasm. *Genome* 49:1510–1527
- Reyna N, Sneller CH (2001) Evaluation of marker-assisted introgression of yield QTL alleles into adapted soybean. *Crop Sci* 41:1317–1321
- Sebolt AM, Shoemaker RC, Diers BW (2000) Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. *Crop Sci* 40:1438–1444
- Shi A, Chen P, Zhang B, Hou A (2010) Genetic diversity and association analysis of protein and oil content in food-grade soybeans from Asia and the United States. *Plant Breeding* 129:250–256
- Smalley MD, Fehr WR, Cianzio SR, Han F, Sebastian SA, Streit LG (2004) Quantitative trait loci for soybean seed yield in elite and plant introduction germplasm. *Crop Sci* 44:436–442
- Song Q, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, Specht JE, Cregan PB (2004) A new integrated genetic linkage map of the soybean. *Theor Appl Genet* 109:122–128
- Song Q, Jia G, Zhu Y, Grant D, Nelson RT, Hwang E-Y, Hyten DL, Cregan PB (2010) Abundance of SSR motifs and development

- of candidate polymorphic markers (BARCSOYSSR_1.0) in soybean. *Crop Sci* 50:1950–1960
- Specht JE, Chase K, Macrander M, Graef GL, Chung J, Markwell JP, Germann M, Orf JH, Lark KG (2001) Soybean response to water: a QTL analysis of drought tolerance. *Crop Sci* 41:493–509
- Tucker DM, Saghai-Maroo MA, Mideros S, Skoneczka JA, Nabati DA, Buss GR, Hoeschele I, Tyler BM, St. Martin SK, Dorrance AE (2010) Mapping quantitative trait loci for partial resistance to *Phytophthora sojae* in a soybean interspecific cross. *Crop Sci* 50:628–635
- Wang D, Graef GL, Procopiuk AM, Deirs BW (2004) Identification of putative QTL that underlie yield in interspecific soybean back-cross populations. *Theor Appl Genet* 108:458–467
- Wang S, Basten CJ, Zeng ZB (2005) Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh
- Wang J, Li H, Zhang L, Li C, Meng L (2011) QTL IciMapping V3.1. Institute of Crop science and CIMMT China, Chinese Academy of Agriculture Sciences, Beijing, China
- Wang X, Jiang G-L, Green M, Scott RA, Hyten DL, Cregan PB (2012) Quantitative trait locus analysis of saturated fatty acids in a population of recombinant inbred lines of soybean. *Mol Breed* 30:1163–1179
- Wang X, Jiang G-L, Green M, Scott RA, Hyten DL, Cregan PB (2014) QTL analysis of unsaturated fatty acids in a recombinant inbred population of soybean. *Mol Breed* 33:281–296
- Wilson RF (2004) Seed composition. In: Boerma RH, Specht JE (eds) Soybeans: improvement, production, and uses, 3rd edn. American Society of Agronomy, Madison, pp 621–669
- Yesudas CR, Bashir R, Geisler MB, Lightfoot DA (2013) Identification of germplasm with stacked QTL underlying seed traits in an inbred soybean population from cultivars Essex and Forrest. *Mol Breed* 31:693–703
- Yuan JZ, Njiti VN, Meksem K, Iqbal MJ, Triwitayakorn K, Kassem MA, Davis GT, Schmidt ME, Lightfoot DA (2002) Identification of yield loci in soybean populations that segregate for disease resistance. *Crop Sci* 42:271–277