Theor Appl Genet (2013) 126:425–433
DOI 10.1007/s00122-012-1990-8

ORIGINAL PAPER

Identification of the quantitative trait loci (QTL) underlying water soluble protein content in soybean

Weiguo Lu · Zixiang Wen · Haichao Li · Daohua Yuan · Jinying Li · Hui Zhang · Zhongwen Huang · Shiyou Cui · Weiijun Du

Received: 10 January 2011 / Accepted: 21 September 2012 / Published online: 7 October 2012

Abstract Water soluble protein content (SPC) plays an important role in the functional efficacy of protein in food products. Therefore, for the identification of quantitative trait loci (QTL) associated with SPC, 212 $F_{2:9}$ lines of the recombinant inbred line (RIL) population derived from the cross of ZDD09454 × Yudou12 were grown along with the parents, in six different environments (location × year) to determine inheritance and map solubility-related genes. A linkage map comprising of 301 SSR markers covering 3,576.81 cM was constructed in the RIL population. Seed SPC was quantified with a macro-Kjeldahl procedure in samples collected over multiple years from three locations (Nantong in 2007 and 2008, Zhengzhou in 2007 and 2008, and Xinxiang in 2008 and 2009). SPC demonstrated transgressive segregation, indicating a complementary genetic structure between the parents. Eleven putative QTL were associated with SPC explaining 4.5–18.2 % of the observed phenotypic variation across the 6 year/location environments. Among these, two QTL (*qsp8-4, qsp8-5*) near GMENOD2B and Sat_215 showed an association with SPC in multiple environments, suggesting that they were key QTL related to protein solubility. The QTL × environment interaction demonstrated the complex genetic mechanism of SPC. These SPC-associated QTL and linked markers in soybean will provide important information that can be utilized by breeders to improve the functional quality of soybean varieties.

Communicated by H. T. Nguyen.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-012-1990-8) contains supplementary material, which is available to authorized users.

W. Lu (✉) · Z. Wen · H. Li · D. Yuan · J. Li · H. Zhang
Institute of Industrial Crops, Zhengzhou National Subcenter for Soybean Improvement/Key Laboratory of Oil Crops in Huanghuaihai Plains, Ministry of Agriculture, P R China, Henan Academy of Agricultural Sciences, Zhengzhou, Henan 450002, China
e-mail: 123bean@163.com

Z. Huang
Henan Institute of Science and Technology, Xinxiang, Henan 453003, China

S. Cui
Jiangsu Yanjiang Institute of Agricultural Sciences, Nantong, Jiangsu 226541, China

W. Du
Faculty of Agronomy, Shanxi Agricultural University, Taigu, Shanxi 030801, China

Introduction

Proteins derived from soybean (*Glycine max* [L.] Merr.) are widely used in food as functional and nutritional ingredients, as well as a substitute for animal-derived proteins (Qi et al. 1997; Rhee 1994). Soy protein can be fractionated into four main groups, on the basis of sedimentation velocity, as 2S, 7S, 11S, and 15S (Rhee 1994). The most abundant soy proteins (70 %) are the globulins, which are largely glycinin (11S) and β-conglycinin (7S), both of which exhibit high solubility (Thanh and Shibasaki 1976). While elevated solubility is not a requirement for all functional properties of food proteins, it does enhance their efficacy (Zayas 1997). The percentage of extractable soluble protein is a traditional criterion in the evaluation of soybean product quality (Malhotra and Coupland 2004; Walsh et al. 2003; Wolf 1969). The utility of soy proteins

in foodstuffs of moderate acidity, such as citric beverages and dressings, may be lessened when solubility is low, especially when the desired functional properties (e.g., foaming and emulsifying) are directly linked to solubility (Walstra 1989). The solubility characteristic is of interest from a scientific as well as economic perspective, since the first step in all soy protein isolation methods is the dissolution of soy flour in water. Any procedures to enhance the content of soluble protein are also likely to improve the final yield of extracted protein.

Current evidence indicates that aromatic amino acids may play important role in protein solubility, due to their high hydrophobicity, calculated from the free energy of transfer (Zayas 1997). However, there is a paucity of data on the genetic drivers of soy protein solubility. Yaklich (2001) analyzed the composition of storage proteins in high-protein soybean and found that the levels of β-conglycinin (7S) and glycinin (11S) in high seed protein lines were elevated above those in normal protein lines. Panthee et al. (2004) reported that three quantitative trait loci (QTL) for glycinin were distributed on chromosomes 17, 19 and 20, while two QTL for β-conglycinin were distributed on chromosomes 16 and 17. In addition, seven genes (*Gy1–Gy7*) had been identified as regulating the production of glycinin (Beilinson et al. 2002; Nielsen et al. 1989), and three genes were determined to control β-conglycinin production in soybean (Harada et al. 1989). So it can be inferred that quantitative differences exist in the soluble fraction of storage protein, which may be initially caused by genetic variations of soybean. Meng et al. (2008) also found substantial variability in the soluble protein content (SPC) among soybean germplasms and mutants. Given these phenotypic and genotypic variations, it will be helpful to isolate markers associated with the QTL that govern soluble properties, thus allowing geneticists to identify individuals that are likely to promote changes in the genome that would enhance functional solubility. A lot of QTL associated to protein content have been mapped in soybean (Chung et al. 2003; Diers et al. 1992; Liu et al. 2009; Nichols et al. 2006; Sebolt et al. 2000; Tajuddin et al. 2003). In all, 86 QTL related to soybean protein content have been reported in Soybase (http://www.soybase.org/). These QTL extend to nearly every chromosome of the soybean genome. However, there appear to be no current reports that focus on mapping QTL associated with soluble protein in soybean.

Detection of QTL that govern soluble protein may provide marker-assisted selection (MAS) opportunities for future improvement of SPC. In this study, our research objective was to identify QTL for SPC in a recombinant inbred line (RIL) population.

Materials and methods

Plant materials

The RIL population used in this study was derived from the cross between ZDD09454, a spring soybean landrace from Shanxi Province with a black seed coat, white flower and higher SPC, and Yudou12, an elite released cultivar in Henan Province with a yellow seed coat, purple flowers and lower SPC. ZDD09454 was used as female parent in order to verify the true hybridity of F_1 plants. To provide each plant enough growing space, all the 11 F_1 plants were planted in two 4 m long rows with 0.8 m interval in the summer of 2002. More than 300 seeds from one selected F_1 plant of the cross ZDD09454 × Yudou12 were harvested to obtain a F_2 population. In the winter of 2002 in Sanya, Hainan Province, China, all these seeds were sowed in 0.4 m row interval and 0.4 m plant distance to ensure each plant produced enough seeds for the next generation. Two hundred and twelve F_2 plants with more than 50 seeds were selected to derive RILs by single seed descent from 2003 to 2006. From 2003 to 2005, the population was planted in Zhengzhou in summer and in Sanya in winter. For each generation, more than five seeds were harvested from each plant to avoid line deletion in case of poor germination.

Field experiment

The RIL population and the parents were planted in a randomized complete block design with three replications at three separate locations: (1) Nantong Experimental Station, Jiangsu Yanjiang Institute of Agricultural Sciences in 2007 and 2008, (2) Zhengzhou Experimental Station, Institute of Industrial Crops, Henan Academy of Agricultural Sciences in 2007 and 2008, and (3) at Xinxiang Experimental Station, Henan Institute of Science and Technology in 2008 and 2009, respectively. Each plot consisted of two, 4 m long rows with 0.4–0.5 m between rows. Planting density was 25 seed/m of row, thinned to 31 plants per row just before the V2 stage. Conventional tillage practices were followed to maintain a weed-free field. No fertilizer was used. Seeds from 20 plants of each plot were harvested and used to measure SPC, protein and oil content.

Measurement of water soluble protein

The seeds from each accession were milled with a sample mill (FW-100, Huanghua, China) for 25 s. 5 g of the milled powder were weighed into a 50 ml plastic tube, followed by 50 g of deionized water. Samples were stirred on an

orbital stirrer (HZP-250, Jinghong, China) at a speed setting of 3 for 1 h at 20 °C. The suspended homogenate was centrifuged at 10,000×*g* for 10 min at 20 °C. The supernatant was decanted from the tube and placed in a 250 ml glass bottle. The pellet remaining at the bottom of the tube was resuspended in 50 ml deionized water and centrifuged again at 10,000×*g* for 10 min at 20 °C. The supernatant was decanted into the same 250 ml bottle containing the original supernatant. The precipitate was resuspended in 50 ml deionized water, centrifuged as described above and twice-washed with deionized water. The washings and the previous supernatant were combined and brought to 250 ml with deionized water in the original 250 ml bottle. Soluble nitrogen in the total supernatant was determined using the macro-Kjeldahl procedure [IDF (International Dairy Federation) 1993]. Crude protein content was estimated by multiplying total nitrogen (N) by 6.25. Protein solubility was expressed as grams of soluble protein/100 g of sample. Seed total protein and oil content were measured using a near infrared spectrophotometer (NIRS) (NIRS 5000, FOSS in North America, Eden Prairie, MN, USA). Diagnostics and calibration of the NIRS at room temperature (~20 °C) occurred daily to ensure accuracy; the instrument was left on during the entire analysis period. Winisi II 1.5 software was used to quantify the spectroscopic scans.

Simple sequence repeat genotyping

Genomic DNA of the parents and each RIL were extracted from freeze-dried leaf tissue according to the method of Doyle and Doyle (1990). To achieve a reasonable coverage of all 20 chromosomes, 1364 SSR markers of BARCSOYSSR_1.0 developed by Song et al. (2004, 2010) were chosen and evaluated for their polymorphism between ZDD09454 and Yudou12. The primer sequences with their linkage group locations are available at http://bldg6.arsusda.gov/cregan/soymap.htm. A total of 361 SSR markers were identified as polymorphic markers and these markers were subsequently utilized for genotyping of the RILs and QTL analysis.

The polymerase chain reaction (PCR) amplification buffer (10 µl) consisted of 20 ng genomic DNA, 0.4 µM forward and reverse primers, 200 µM of each dNTPs, 1× PCR buffer (10 mM of Tris-HCl, pH 8.3, 50 mM of KCl), 2 mM of $MgCl_2$ and 0.5 U of Taq DNA polymerase. PCR was programmed with an initial denaturing at 94 °C for 3 min, followed by 30 cycles of 94 °C for 30 s, 55 °C for 60 s and 72 °C for 1 min, with a final extension at 72 °C for 8 min. PCR amplification was performed with an Eppendorf AG 2231 cycler (Eppendorf, Hamburg, Germany). PCR products (2.5 µl/lane) were separated on 8 % polyacrylamide gel and polymorphisms were detected after silver staining.

Construction of the genetic map

MAPMAKER/EXP 3.0 (Lander et al. 1987) was used to analyze the linkage between markers. Markers were grouped at a likelihood of odds (LOD) ratio of 3.0 and ordered at an LOD of 2.0. A minimum (LOD) of ≥3.0 and a maximum distance of ≤50 cM were used to test linkages among markers. Chromosome numbers were assigned to the linkage groups according to the map of Song et al. (2004, 2010).

Data analysis and QTL mapping

Analysis of data was conducted using SAS 8.0 statistical software (SAS Institute, Cary, NC, USA). Arithmetic mean values of the three replicates were calculated for each variable for RILs, and used for QTL mapping. The heritability of the trait in each location in the population was estimated on an entry mean basis as follows:

$$h^2 = \sigma_g^2/(\sigma_g^2 + \sigma_{ge}^2/n + \sigma_e^2/nr), \quad (1)$$

where h^2 is narrow-sense heritability, σ_g^2 is genotypic variance, σ_{ge}^2 is genotype × environment variance, σ_g^2 is error variance, n is number of environments and r is the number of replications (Nyquist 1991).

The means of each RIL from each environment were used for QTL analysis, using Cartographer version 2.5 (Wang et al. 2005). Composite interval mapping (CIM) (model 6) was used to map QTL. The genome was scanned at 2 cM intervals and the window size was set at 10 cM. Cofactors were chosen using the forward and backward regression method at $P(\text{Fin}) = P(\text{Fout}) = 0.05$. For each environment, two or more peaks on one chromosome were considered as different QTL when they were separated by a minimum distance of 20 cM, otherwise the highest peak was chosen to more closely approximate the position of the QTL (Ravi et al. 2011). A joint analysis of the phenotypic data for all environments was used to determine the QTL by environment (Q × E) interaction (Jiang and Zeng 1995).

For Q × E interaction analysis, LOD thresholds of 3.27 were used. The LOD scores for declaring a significant QTL were 2.53 and 6.26 for the single-trait and joint QTL analysis, respectively, by permutation test analyses (1,000 permutations, significance level = 5 %) as described previously (Churchill and Doerge 1994). The nomenclature of the QTL included four parts, beginning with '*q*' followed by an abbreviation of the trait name (*sp*), chromosome number and the number of QTL affecting the trait, following the recommendations of the Soybean Germplasm Coordination Committee.

Results

Polymorphism and linkage

In total, 361 SSR markers were polymorphic between the two parents and these markers were used to construct the genetic linkage group map. Finally, 301 markers formed 26 linkage groups, in which chromosome 1, 5, 13 and 18 were divided into two minor linkage groups and chromosome 2 into three segments (supplementary material 1). On average there were 8.3 markers per group, ranging from 5 (chromosome 16) to 15 (chromosome 8) markers. The total map encompassed 3,576.81 cM with an average distance of 13.0 cM between markers. In general, the mapping orders of the SSR markers detected in this study were in good accordance with the integrated soybean genetic linkage map (Song et al. 2004, 2010).

Soluble protein phenotypic variation and correlations

There was considerable variability in the seed SPC among the 212 RILs across six different environments. The RIL population demonstrated prominent negative and positive transgressive segregation for the SPC (Fig. 1). For example, the SPC in the RILs displayed a continuous frequency distribution ranging from a low of 12.25 % to a high of 31.26 % at Nantong in 2007, and a low of 24.27 % to a high of 37.25 % at Xinxiang in 2008. The average SPC of ZDD09454 (30.55 %) and Yudou12 (25.82 %) were significantly different. Analysis of variance (ANOVA) indicated a significant ($P < 0.001$) genetic variation for SPC (Table 1) across three locations. While in Nantong and Xinxiang, the lines × year interaction (σ_g^2) were also large (Table 1). Therefore, both the genetic effects and the genotype × environment epistatic should be considered in the proceeding mapping program.

Narrow-sense heritability was moderate for the trait ranging from 0.399 to 0.741 among the three different locations, which indicated that SPC could be influenced, to a large extent, by the environment in which plants were grown. In certain places, like Zhengzhou, the genetic effect was the main component for phenotype variation.

In this study, SPC was significantly correlated to protein content ($r = 0.386$, $P < 0.01$) and to the SPC: protein content ratio ($r = 0.905$, $P < 0.01$) (Table 2). There was no significant correlation between soluble protein and oil ($r = 0.034$, $P = 0.31$).

Confirmed QTL for soluble protein content

Eleven putative QTL were found to be associated with the SPC in six environments. These QTL were scattered across 10 soybean chromosomes, including 5, 6, 7, 8, 9, 11, 12, 18, 19 and 20 (Table 3; Fig. 2). The detected QTL individually accounted for 4.5–18.2 % of the phenotypic variation of SPC with a LOD range of 2.53–9.69. The additive effects of the detected QTL were estimated from CIM results as the mean effect of replacing both Yudou12 alleles with ZDD09454 alleles at the locus studied. Thus, a positive effect implied a higher value for the trait conferred by the ZDD09454 allele, and vice versa.

By comparing the detected QTL (Table 3; Fig. 2), a major genomic region with high LOD scores and R^2 was identified as being located around the marker interval of Sat_215-GMENOD2B on chromosome 8. Proximal to the interval, two QTL (*qsp8-4* and *qsp8-5*) were detected for SPC across more than four of the experimental environments. At these loci, ZDD09454 alleles were positive for

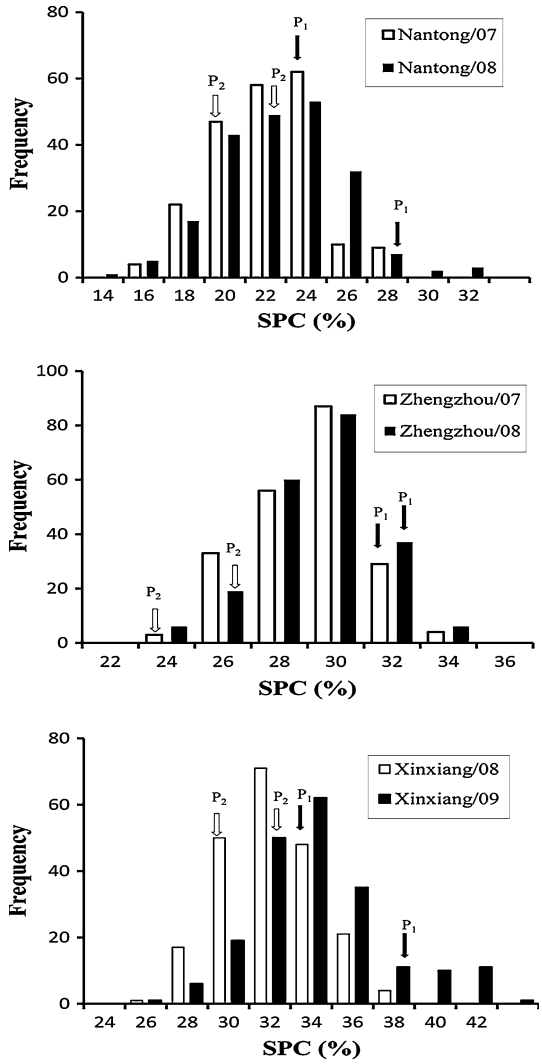

Fig. 1 Frequency distribution of the percentage of soluble protein content (%) among the $F_{2:9}$ RILs from the cross of ZDD09454 × Yudou12 in six environments. The means of the SPC values among different environments of the RIL population are shown. P_1 ZDD09454, P_2 Yudou 12

Table 1 General characteristics for water soluble protein content of soybean RILs and parents

Location	Years	Parents		RIL population		F value		σ^2_g	σ^2_{ge}	σ^2_e	h^2
		Yudou12[a]	ZDD09454[a]	Mean ± SD[a]	Range[a]	RIL	RIL × E				
Nantong	2007	20.42	24.15	21.63 ± 3.02	12.25–31.26	1.79***	2.06***	25.910	16.349	7.921	0.401
	2008	22.78	27.17	21.04 ± 2.44	15.55–27.54	3.80***					
Zhengzhou	2007	24.01	30.42	28.40 ± 2.02	22.56–32.96	2.14***	0.94[NS]	11.912	4.954	5.300	0.741
	2008	26.56	31.54	28.15 ± 2.00	22.49–33.04	2.11***					
Xinxiang	2008	29.39	33.04	31.22 ± 2.28	24.27–37.25	2.18***	3.69***	36.032	22.049	5.973	0.399
	2009	31.79	37.01	33.62 ± 3.85	24.75–43.41	3.40***					

NS not significant; h^2 heritability on an entry mean basis

*** Significance at $P < 0.001$

[a] These values are percent (%)

Table 2 Pearson correlation coefficients between soluble protein (SP), protein (P), and the SP:P ratio and oil content in the RIL population

Traits	Soluble protein	Protein	SP:P
Protein	0.386**		
SP:P	0.905**	−0.042[ns]	
Oil	0.034 [ns]	−0.734**	0.378**

ns nonsignificant

** Significance at the $p < 0.01$; * significance at $p < 0.05$ level

the trait. The high-SPC parent ZDD09454 alleles were positive for SPC at 8 loci among all of the 11 QTL. The additive of the remaining three QTL was negative, indicating that the low-SPC genotype carried the favorable allele for this trait. Three QTL (*qsp18-9, qsp11-7* and *qsp20-11*) were located in the intervals of Sat_204-Sat_094 on chromosome 18, Sat_156-11_0203 on chromosome 11 and AB002807-Satt270 on chromosome 20, respectively.

The LOD scores for the joint analysis and Q × E interaction (Table 3) indicated that *qsp8-4* and *qsp8-5* were generally stable across different environments, with Q × E interaction LOD scores of less than 3.27; those found on other genomic regions were mainly expressed in specific environment.

Discussion

Protein solubility is a critical factor in both food quality and aesthetic appearance of soybean products. Low solubility will cause proteins to separate and settle out, resulting in irregular and reduced dispersion in food products. High-protein solubility is required to obtain preferable emulsifying and foaming properties. For example, to form a gel, a protein must first form a well-dispersed mixture in the solvent, followed by limited protein–protein interactions, yielding a solid material with elastic properties (Malhotra and Coupland 2004). If an adequate solution is not formed, i.e., the protein–protein interactions are much stronger than the combination of protein–solvent interactions and mixing energy, a strong gel will not be produced (Kinsella 1979). Conversely, soluble protein is easily dissolved in the rumen, and therefore, immediately available to rumen microbes which convert it quickly to ammonia when used in animal feeds. Past efforts to improve SPC in soybean have included enzymatic and other chemical approaches. Malhotra and Coupland (2004) reported that anionic surfactants were effective in solubilizing soybean seed proteins. Bromelain digestion has also proven to be effective for improving soybean protein solubility (Ortiza and Wagner 2002). Walsh et al. (2003) reported that hydrolysates, hydrolysates of cross-linked soybean protein isolate (SPI) and the cross-linked products of SPI hydrolysates all displayed significant ($P < 0.05$) increases in solubility compared to unmodified SPI. However, the processing of soy protein in this manner leads to a more expensive finished product, and is therefore not a sustainable approach. Unfortunately, soybean soluble protein content has been difficult to evaluate by phenotype. Increasing the genotype selection intensity by MAS, on the other hand, will allow development of cultivars with enhanced levels of soluble protein.

To date, little data are available regarding the genetic control and molecular-physiological mechanisms contributing to elevated SPC in soybean. Most of the QTL studies conducted so far have focused on protein content with little attention given to its components. We identified only a few published studies of QTL mapping for essential amino acids in soybean (Panthee et al. 2006), and the 11S and 7S fractions (Liu et al. 2009; Panthee et al. 2004; Tsubokura et al. 2006). In fact, there is very limited information on QTL mapping for protein components (or subunits) in any species. Wang et al. (2001a) mapped QTL involved in determining the free amino acid content in *Zea mays*. Wang et al. (2001b) also found two QTL associated with an

Table 3 Position and effects of QTL detected to be significantly ($\alpha = 5$ %) associated with soybean soluble protein content in the ZDD09454 × Yudou12 $F_{2:9}$ RIL population

QTL	Chr.	Marker interval	Position (cM)	LOD			Additive	R^2 (%)
				Single Environment	Joint	Q × E		
qsp5-1	5	05_1318-05_1370	72.50	4.97 (Zhengzhou/08)	5.48	0.67	0.634	9.7
qsp6-2	6	Sat_213-GMAC7L	147.70	2.79 (XinXiang/08)	3.49	3.19	0.690	9.1
qsp7-3	7	Satt336-Satt308	285.80	4.63 (Nantong/07)	4.88	4.09	0.718	8.3
qsp8-4[a]	8	Sat_115-GMENOD2B	141.00	4.06 (Nantong/07)	12.76	2.41	0.874	12.1
			145.20	5.81 (Zhengzhou/08)			0.719	12.1
			143.10	6.87 (Zhengzhou/07)			0.826	15.8
			139.30	5.8 (XinXiang/09)			1.602	16.6
			144.90	8.45 (XinXiang/08)			0.994	18.2
qsp8-5[a]	8	GMENOD2B- Sat_215	152.80	4.09 (Nantong/07)	13.34	2.44	0.725	8.1
			150.80	5.94 (Zhengzhou/08)			0.670	10.4
			150.80	6.49 (Zhengzhou/07)			0.713	11.3
			152.80	9.69 (XinXiang/08)			0.985	17.9
qsp9-6	9	Satt475-09_1155	14.60	2.53 (XinXiang/08)	4.03	3.53	0.685	8.8
qsp11-7	11	Sat_156-11_0203	59.90	2.64 (Nantong/07)	2.92	1.55	−0.486	5.7
qsp12-8	12	Satt541-Satt469	78.01	2.53 (Nantong/08)	3.86	1.42	0.651	4.5
qsp18-9	18	Sat_204-Sat_094	142.50	4.61 (Zhengzhou/08)	7.11	3.78	−0.548	7.2
			141.70	4.89 (Zhengzhou/07)			−0.782	15.1
qsp19-10	19	Satt143-Sat_195	76.90	4.41 (Nantong/08)	4.01	3.20	0.747	5.9
qsp20-11	20	AB002807-Satt270	49.31	2.97 (Zhengzhou/07)	4.51	1.38	−0.713	12.6

Chr. chromosome number of soybean; *LOD* (Log10 of the likelihood odds ratio) the probability associated with the most likely location of the detected QTL; *Marker interval* the interval within which QTL were mapped; *Position* the position of the peak of the QTL on chromosome; *Joint* LOD score in the joint analysis of six environments; *Q* × *E* LOD score value for QTL–environment interaction in the joint analysis of six environments

[a] For these two QTL, we list all the LOD peaks in each environment no matter how close they linked in order to show their environmental stability

elongation factor involved in lysine biosynthesis in *Z. mays*. Loudelt et al. (2003) mapped nine QTL for amino acids in *Arabidopsis*.

Traits that show substantial correlation may have loci in common, and are thus manifested through linkage or pleiotropy (Aastveit and Aastveit 1993). Therefore, QTL associated with SPC are expected to be associated with protein. In the present study, we identified eight SPC-related QTL that were located at, or closely linked to, previously-identified QTL detected for soybean protein or its components. For example, Nichols et al. (2006) reported a fine-mapped protein QTL located at intervals Satt239-ACG9b from chromosome 20, which just overlapped the intervals of *qsp20-11* detected in the present study. A QTL associated with the glycinin fraction of storage protein detected between Satt564 and Satt191 from chromosome 18 (Panthee et al. 2004) was found to be located at the intervals of *qsp18-9* in the present study. Two protein QTL (*prot1-7* and *prot21-11*), linked to molecular markers A023_1 and Satt308, were reported from chromosome 19 (Brummer et al. 1997) and chromosome 7 (Kabelka et al. 2004), respectively. These QTL also overlapped the marker intervals of *qsp19-10* and *qsp7-3*, respectively, in the present study. In addition, four other protein-related QTL, namely *prot3-11*(Brummer et al. 1997), *prot17-4* (Tajuddin et al. 2003), *prot9-1*(Orf et al. 1999) and *prot5-3* (Lee et al. 1996) were found, in the current investigation, to be closely linked to *qsp12-8*, *qsp8-5*, *qsp5-1* and *qsp20-11*, respectively. Higher SPC are likely to improve the final yield of extract protein, therefore, desirable output of soybean production, like Chinese Tofu and soymilk, will be obtained from soybean with higher SPC. Zhang et al. (2008) reported that two QTL *qODT*-A2-1 and *qODS*-A2-1 located between Satt424 and Sat_162 on chromosome 8 (LG A2). The marker interval is just near the loci *qsp8-5*. *qODT*-A2-1, associated with dried tofu output, could explain 15.7–28.2 % of the total phenotypic variation. *qODS*-A2-1, related to dried soymilk output, could explain 30.0–34.8 % of the phenotypic variation.

Not all QTL associated with SPC were the same as those determined to be associated with seed protein content. We found only relatively low correlation ($r = 0.386$) between SPC and total protein content. Therefore, even though

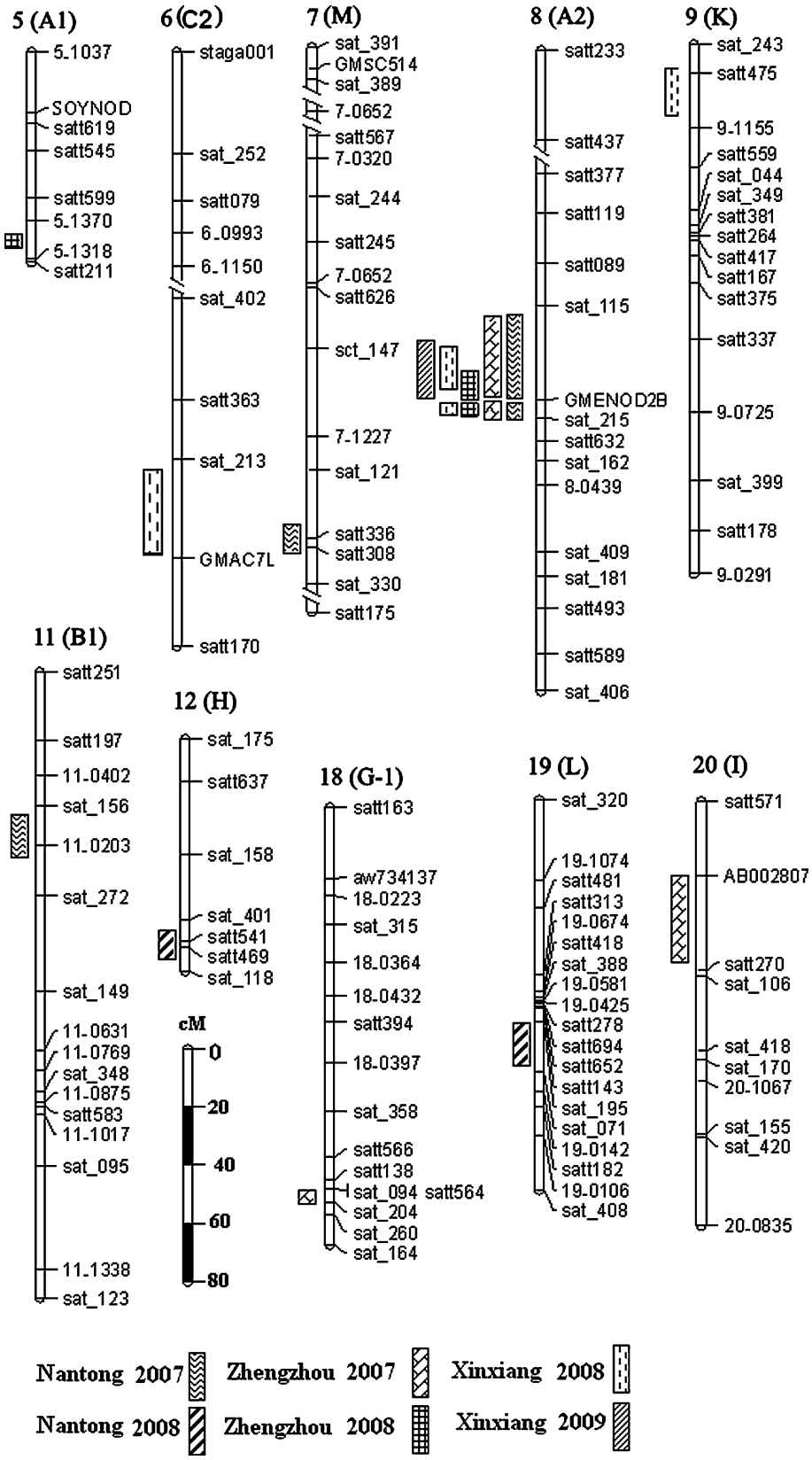

Fig. 2 Summary of QTL locations detected in this study. The *symbol in the parentheses* indicates the corresponding linkage group. QTL represented by *bars*, are shown to the *left* of the chromosomes, close to their corresponding markers. A reduction of one LOD value from the LOD peaks is used to define borders of the confidence intervals. Chromosomes with no QTL are omitted from this figure

soluble protein is a significant protein component, the manner in which SPC is governed in genomic regions differs from the processes that affect the total seed protein content. These results also demonstrated that the QTL × environment interaction was an important property of many QTL underlying SPC, as evidenced by the presence of three QTL with significant Q × E LOD scores (Table 3). Most of the detected QTL expressed only in one environment, which might explain the moderate heritablility of SPC. Result of ANOVA also showed that the variances of genetic × environment were higher in some places. Therefore, the QTL × environment interaction should not be ignored if MAS is applied for the manipulation of SPC.

We had tried to evaluate the Q × Q epistatic effects in the mapping population with the program QTL Network-v2.0 (Yang et al. 2007; Yang et al. 2008). But only one significant additive epistatic interaction was detected with the estimated R^2 of 3.85 % (data not shown). It implied that Q × Q epistatic effects was not the main genetic effect in this population. No QTL × environment interaction had been detected by this program. It might be that the larger population and more marker information are need for detecting epistatic effect.

Based on the analysis of the additive effects of the detected QTL, we concluded that an increase in SPC was not only due to the allelic contribution of the high-SPC parent ZDD09454, but also to the allelic contribution of the low-SPC parent Yudou12. The results suggested that, accumulation of favorable alleles in the progeny were possible from both of the parents.

Two QTL, *qsp8-4* and *qsp8-5*, linked closely on chromosome 8. Based on the limited information of this study, we couldn't determine whether they were two clustering loci or one locus with larger genetic effect. More markers on this chromosome are needed in the further study, especially in the large gap between Sat_115 and GMENOD2B. But for MAS program, the closely linked markers GMENOD2B and Sat_215 will be helpful, because the genetic effects on SPC of these two QTL were larger and stable across different environments (mostly $R^2 > 10$ %, Table 3). The favorable alleles of the two QTL were all from the same parents ZDD09454. Therefore, selecting with the ZDD09454 alleles of the two markers will be beneficial in the development of higher SPC soybean varieties. Although the remaining QTL for SPC reported here were not as great as the two QTL previously mentioned, they may nevertheless yield substantial initial data that could be utilized in an effective MAS program to achieve genetic gains.

Acknowledgments The germplasm ZDD09454 was kindly provided by Prof. Chang Ru-zhen and Doctor Qiu Li-juan, Institute of crop sciences, Chinese Academy of Agricultural Sciences. The project was supported by the Natural Science Foundation of China (30971818, 30771360), Zhengzhou technological leading training program (096SYJH14103), Innovation Scientists and Technicians Troop Construction Projects of Henan Province (094100510004) and Doctor Foundation of Henan Institute of Science and Technology (2008005).

References

Aastveit AH, Aastveit K (1993) Effects of genotype–environment interactions on genetic correlations. Theor Appl Genet 86: 1007–1013

Beilinson V, Chen Z, Shoemaker RC, Fischer RL, Goldberg RB, Nielsen NC (2002) Genomic organization of glycinin genes in soybean. Theor Appl Genet 104:1132–1140

Brummer EC, Greaf GL, Orf JH, Wilcox JR, Shoemaker RC (1997) Mapping QTL for seed protein and oil content in eight soybean populations. Crop Sci 37:370–378

Chung J, Babka HL, Greaf 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

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963–971

Diers BW, Keim P, Fehr WR, Shoemaker RC (1992) RFLP analysis of soybean seed protein and oil content. Theor Appl Genet 83:608–612

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15

Harada JJ, Barker SJ, Goldberg RB (1989) Soybean beta-conglycinin genes are clustered in several DNA regions and are regulated by transcriptional and posttranscriptional processes. Plant Cell 1:415–425

IDF (International Dairy Federation) (1993) Milk: determination of nitrogen content (Kjeldahl method), (standard 20B, part 1) Brussels

Jiang C, Zeng ZB (1995) Multiple trait analysis of genetic mapping for quantitative trait loci. Genetics 140:1111–1127

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 introduction. Crop Sci 44:784–791

Kinsella JE (1979) Functional properties of soy proteins. J Am Oil Chem Soc 56:242–258

Lander ES, Green P, Abrahamson J, Barlow A, Daly M, Lincoln S, Newburg L (1987) Mapmaker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Lee SH, Bailey MA, Mian MAR, Carter TE Jr, 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

Liu SH, Zhou RB, Yu DY, Chen SY, Gai JY (2009) QTL mapping of protein related traits in soybean [*Glycine max* (L.) Merr.]. Acta Agron Sini 12:2139–2149

Loudelt O, Chaillou S, Merigout P, Talbotec J, Daniel-Vedele F (2003) Quantitative trait loci analysis of nitrogen use efficiency in *Arabidopsis*. Plant Physiol 131:345–358

Malhotra A, Coupland JN (2004) The effect of surfactants on the solubility, zeta potential, and viscosity of soy protein isolates. Food Hydrocolloids 18:101–108

Meng T, Meng FW, Zhang D, Hao ZB, Li ZY (2008) Analysis of some quality characteristics in soybean mutants. Guihaia 3:382–385

Nichols DM, Glover 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

Nielsen NC, Dickinson CD, Cho TJ, Thanh VH, Scallon BJ, Fischer RL, Sims TL, Drews GN, Goldberg RB (1989) Characterization of the glycinin gene family in soybean. Plant Cell 1:313–328

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

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(6): 1642–1651

Ortiza SEM, Wagner JR (2002) Hydrolysates of native and modified soy protein isolates: structural characteristics, solubility and foaming properties. Food Res Int 35:511–518

Panthee DR, Kwanyuen P, Sams CE, West DR, Saxton AM, Pantalone VR (2004) Quantitative trait loci for beta-conglycinin (7S) and glycinin (11S) fractions of soybean storage protein. J Am Oil Chem Soc 81:1005–1012

Panthee DR, Pantalone VR, Sams CE, Saxton AM, West DR, Orf JH, Killam AS (2006) Quantitative trait loci controlling sulfur containing amino acids, methionine and cysteine, in soybean seeds. Theor Appl Genet 112:546–553

Qi M, Hettiarachchy NS, Kalapathy U (1997) Solubility and emulsifying properties of soy protein isolates modified by pancreatin. J Food Sci 6:110–115

Ravi K, Vadez V, Isobe S, Mir RR, Guo Y, Nigam SN, Gowda MVC, Radhakrishnan T, Bertioli DJ, Knapp SJ, Varshney RK (2011) Identification of several small main-effect QTLs and a large number of epistatic QTLs for drought tolerance related traits in groundnut (*Arachis hypogaea* L.). Theor Appl Genet 122:1119–1132

Rhee KC (1994) Functionality of soy proteins. In: Hettiarchchy NA, Ziegler GR (eds) Protein functionality in food systems. Marcel Dekker, New York, pp 311–324

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

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

Song QJ, Jia GF, Zhu YL, Grant D, Nelson RT, Hwang EY, Hyten DL, Cregan PB (2010) Abundance of SSR motifs and development of candidate polymorphic SSR markers (BARCSOYSSR_1.0) in soybean. Crop Sci 50:1950–1960

Tajuddin T, Watanabe S, Yamanaka N, Harada K (2003) Analysis of quantitative trait loci for protein and lipid contents in soybean seeds using recombinant inbred lines. Breed Sci 2:133–140

Thanh VH, Shibasaki K (1976) Major proteins of soybean seeds. A straight forward fractionation and their characterization. J Agric Food Chem 24:1117–1121

Tsubokura Y, Hajika M, Harada K (2006) Molecular marker associated with beta-conglycinin deficiency in soybean. Breed Sci 56:113–117

Walsh DJ, Cleary D, McCarthy E, Murphy S, FitzGerald RJ (2003) Modification of the nitrogen solubility properties of soy protein isolate following proteolysis and transglutaminase cross-linking. Food Res Int 36:677–683

Walstra P (1989) Principles of foam formation and stability. In: Wilson AJ (ed) Foam: physics, chemistry and structure. Springer, London, pp 1–15

Wang X, Stumpf DK, Larkins BA (2001a) Aspartate kinase 2. A candidate gene of a quantitative trait locus influencing free amino acid content in maize endosperm. Plant Physiol 125: 1778–1787

Wang X, Won Y, Kim CS, Larkins BA (2001b) Quantitative trait locus mapping of loci influencing elongation factor 1alpha content in maize endosperm. Plant Physiol 125:1271–1282

Wang SC, Basten CJ, Zeng ZB (2005) Windows QTL Cartographer v2.5. Department of Statistics, North Carolina State University, Raleigh, USA

Wolf WJ (1969) Chemical and physical properties of soybean proteins. Bakers Digest 43:30–35

Yaklich RW (2001) beta-Conglycinin and glycinin in high-protein soybean seeds. J Agric Food Chem 49:729–735

Yang J, Zhu J, Williams RW (2007) Mapping the genetic architecture of complex traits in experimental populations. Bioinformatics 23:1527–1536

Yang J, Hu CC, Hu H, Yu RD, Xia Z, Ye XZ, Zhu J (2008) QTLNetwork: mapping and visualizing genetic architecture of complex traits in experimental populations. Bioinformatics 24: 721–723

Zayas JF (1997) Solubility of proteins. In: Zayas JF (ed) Functionality of proteins in food. Springer, Berlin, pp 60–75

Zhang HM, Zhou B, Zhao TJ, Xing H, Chen SY, Gai JY (2008) QTL mapping of tofu and soymilk output in RIL population NJRISX of soybean. Acta Agron Sin 34:67–75