



# Chromosome segment detection for seed size and shape traits using an improved population of wild soybean chromosome segment substitution lines

Hongyan Yang<sup>1,2</sup> · Wubin Wang<sup>1,2,3,4</sup> · Qingyuan He<sup>1,2</sup> · Shihua Xiang<sup>1,2</sup> · Dong Tian<sup>1,2</sup> · Tuanjie Zhao<sup>1,2,3,4</sup> · Junyi Gai<sup>1,2,3,4</sup>

Received: 28 February 2017 / Revised: 22 June 2017 / Accepted: 1 September 2017 / Published online: 19 September 2017  
© Prof. H.S. Srivastava Foundation for Science and Society 2017

**Abstract** Size and shape of soybean seeds are closely related to seed yield and market value. Annual wild soybeans have the potential to improve cultivated soybeans, but their inferior seed characteristics should be excluded. To detect quantitative trait loci (QTLs)/segments of seed size and shape traits in annual wild soybean, its chromosome segment substitution lines (CSSLs) derived from *NN1138-2* (recurrent parent, *Glycine max*) and *N24852* (donor parent, *Glycine soja*) and then modified 2 iterations (coded *SojaCSSLP3*) were improved further to contain more lines (diagonal segments) and less heterozygous and missing portions. The new population (*SojaCSSLP4*) composed of 195 CSSLs was evaluated under four environments, and 11, 13, 7, 15 and 14 QTLs/segments were detected for seed length (SL), seed width (SW), seed roundness (SR), seed perimeter (SP) and seed cross section area (SA), respectively, with all 60 wild allele effects negative. Among them, 16 QTLs/segments were shared by 2–5 traits, respectively, but 0–3 segments for each of the 5 traits were independent. The non-shared Satt274 and shared Satt305, Satt540 and Satt239 were major segments,

along with other segments composed of two different but related sets of genetic systems for SR and the other 4 traits, respectively. Compared with the literature, 7 SL, 5 SW and 2 SR QTLs/segments were also detected in cultivated soybeans; allele distinction took place between cultivated and wild soybeans, and also among cultivated parents. The present mapping is understood as macro-segment mapping, the segments may be further dissected into smaller segments as well as corresponding QTLs/genes.

**Keywords** Chromosome segment substitution line (CSSL) · QTL/segment detection · Seed shape · Seed size · Wild soybean (*G. soja* Sieb. and Zucc.)

## Introduction

The cultivated soybean (*Glycine max* (L.) Merr.) was domesticated from annual wild soybean (*Glycine soja* Sieb. and Zucc.) in ancient China (Hymowitz 1990). Compared with the cultivated soybean, the characteristics of wild soybean include a higher protein content, tolerance to biotic and abiotic stresses, floriferousness, more pod number per plant and so on. Introduction of elite wild genes into the cultivated soybean may create new genetic materials with greater breeding value. Seed size and shape traits, including seed length, width and height, are closely related to seed yield (Fatokun et al. 1992), and also affect the morphological quality of marketable seed (Hu et al. 2013). In some special soy food products, such as natto, miso, edamame and tofu, the requirements for size and shape of soybean seed are different. For example, round seed is usually used for edible-type soybean, while larger seeds are preferred for edamame, miso and tofu, and smaller ones for natto, a fermented food (Basra 1995).

✉ Junyi Gai  
sri@njau.edu.cn

<sup>1</sup> Soybean Research Institute, Nanjing Agricultural University, Nanjing 210095, Jiangsu, China

<sup>2</sup> National Center for Soybean Improvement, Ministry of Agriculture, Nanjing 210095, Jiangsu, China

<sup>3</sup> Key Laboratory of Biology and Genetic Improvement of Soybean, Ministry of Agriculture, Nanjing 210095, Jiangsu, China

<sup>4</sup> National Key Laboratory for Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing 210095, Jiangsu, China

Therefore, the soybean traits of seed size and shape are economically relevant. More recently, these traits have become breeding targets dictated by morphological market requirements. As a consequence of long term breeding efforts, the cultivated soybean has many improved traits, such as increased seed weight, seed length, seed width, seed roundness, seed perimeter, seed cross section area and seed oil content; however, seed protein content is somewhat reduced. These traits are controlled by multiple genes with complex interactions. When developing soybean cultivars for human consumption, breeders need to restrain the inferior wild genes/alleles associated with seed size and shape traits when wild soybean is used. This depends on knowledge of the genetic constitution of these wild characteristics.

To-date, 18 quantitative trait loci (QTL) for seed length and 16 QTL for seed width in soybeans have been reported in SoyBase (Moongkanna et al. 2011; Salas et al. 2006). Parts of genes conferring seed size and shape traits have been cloned in rice, tomato, wheat and maize etc. such as *GW2* and *GS3* for rice (Mao et al. 2010; Song et al. 2007), *fw2.2 ovate* and *sun* for tomato (Frary et al. 2000; Liu et al. 2002; Van der Knaap and Tanksley 2001), *Gln1-4* for maize (Martin et al. 2006). However, no results have been reported regarding fine mapping and gene cloning using CSSLs in soybean.

In the literature, QTL mapping for seed size and shape traits was conducted on cultivated soybeans using RIL (recombinant inbred line) populations; few results were associated with wild soybeans (Yang et al. 2011). Recently, the construction of advanced mapping populations, for example near-isogenic lines (NILs) and chromosome segment substitution lines (CSSLs) has received enormous attention (Bernacchi et al. 1998; Zamir and Geiger 2001). To-date, CSSLs have been used in genomic studies of various crops. In tomato plants, a number of fruit size and shape QTLs were detected using CSSL populations (Eshed and Zamir 1995) and some of them were fine mapped (Alpert and Tanksley 1996). It is commonly understood that more QTLs could be

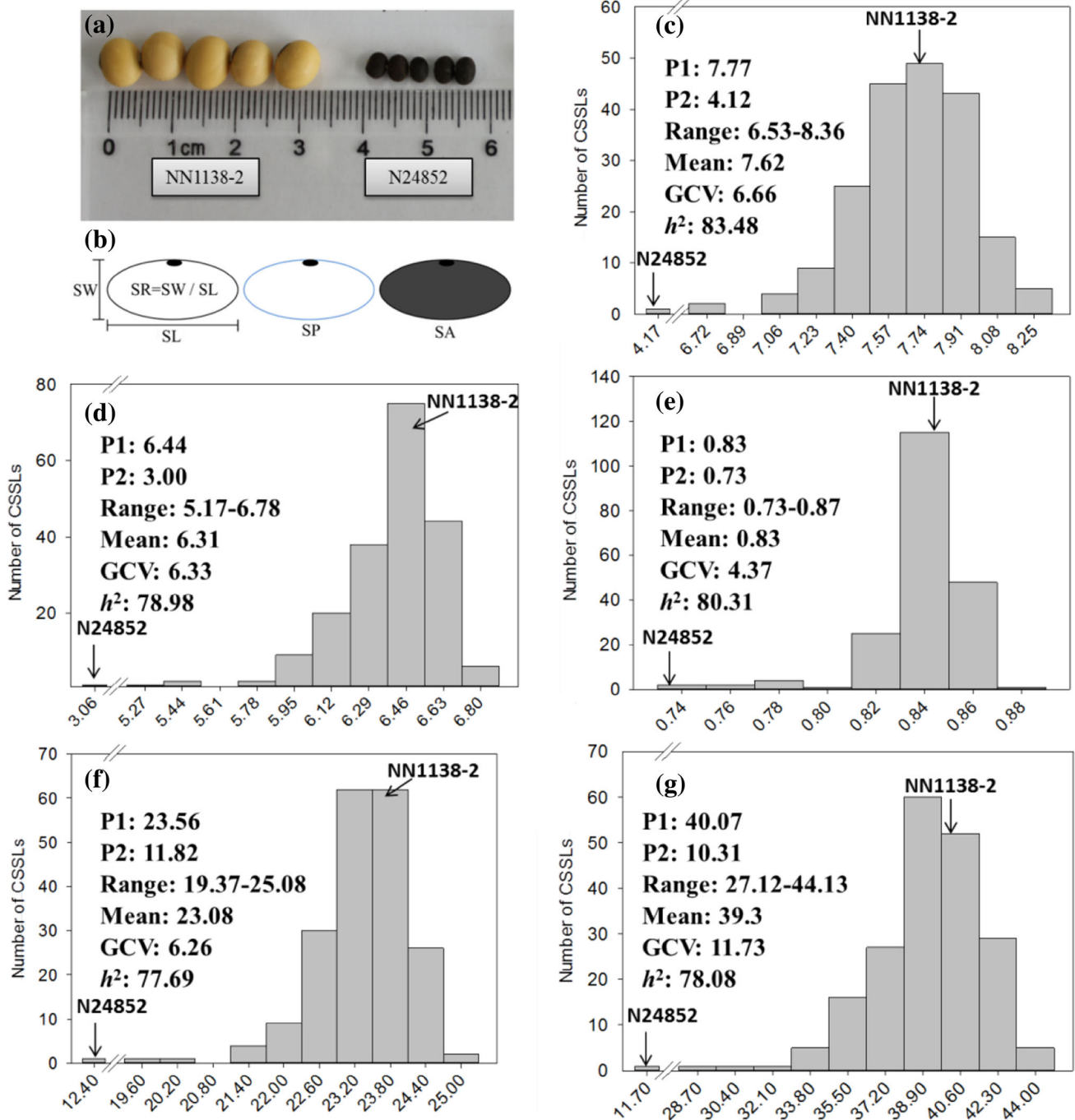
discovered based on advanced mapping populations than on primary mapping populations. Ideally, each line carries a segment from the donor parent, it is easy to detect whether the segment in the line associates with the interested traits by multiple mean comparisons between each line and the recurrent parent following the standard analysis of variance (ANOVA) (Belknap 2003). Since it is time-consuming and laborious to construct a single segment substitution line (SSSL), usually a preliminary CSSL population is used in which each line carries several segments from the donor parent (Wang et al. 2006).

In our research team, Wang et al. (2013a) developed a first population of CSSLs named *SojaCSSLP1* by using the donor parent *N24852* (*G. soja*) and the recurrent parent *NN1138-2* (*G. max*). This CSSL population was improved by Xiang et al. (2015) and He et al. (2014), and designated *SojaCSSLP2* and *SojaCSSLP3*, respectively. By using these CSSL populations, the wild QTLs/segments associated with seed quality traits and other agronomic traits were detected (He et al. 2014; Wang et al. 2012, 2013a, b). In the genome composition graph of these CSSL populations, the segments on the diagonal were continuously connected, but many segments on the two sides of the diagonal remained, unevenly distributed, with some heterozygous segments. These redundant, uneven, heterozygous and missing segments may influence the mapping accuracy, therefore, the CSSL populations were improved with each iteration. However, as such improvements are time-consuming and costly, our aim is to make gradual improvements over time. In the present study, we aimed to improve the *SojaCSSL* populations initially; then subsequently, the QTL/segment mapping of the seed size and shape traits were conducted. Table 1 shows the genome composition of the three previous CSSL populations which indicate that gradual improvements have been achieved. Based on the improvement, the new population was used in the detection of wild QTLs/segments that confer seed size and seed shape traits, including seed length (SL), seed width (SW), seed roundness (SR), seed perimeter (SP) and seed cross section area (SA).

**Table 1** Comparisons of genome composition among four *SojaCSSL* populations

Population	No. of lines	NN1138-2 (%)	N24852 (%)	Heterozygous (%)	Missing (%)	Diagonal segment (cM)	
						Range	Average
<i>SojaCSSLP1</i>	151	95.09	3.68	1.18	0.05	2.69–105.74	34.48
<i>SojaCSSLP2</i>	152	94.66	4.02	0.96	0.35	7.39–72.23	34.97
<i>SojaCSSLP3</i>	158	95.04	3.73	1.19	0.03	3.99–64.60	24.00
<i>SojaCSSLP4</i>	195	95.33	3.67	0.98	0.02	2.22–64.10	21.11

*NN1138-2* (%): the recovered genome portion of the recurrent parent *NN1138-2* in terms of homozygous segment length; *N24852* (%): the donor parent genome portion in terms of homozygous segment length; Heterozygous (%): the genome portion of heterozygous segments in terms of segment length; Missing (%): the missing segment portion in terms of segment length which is not readable for its marker type



**Fig. 1** The frequency distributions of the SL, SW, SR, SP and SA traits in *SojaCSSLP4* along with their parental performances. **a** The pictures of *NN1138-2* and *N24852* seeds, **b** graphic explanation of measurement method of seed traits; **c** frequency distribution of SL (seed length); **d** frequency distribution of SW (seed width);

**e** frequency distribution of SR (seed roundness); **f** frequency distribution of SP (seed perimeter) and **g** frequency distribution of SA (seed cross section area). P1: *NN1138-2*, P2: *N24852*. GCV: coefficient of genetic variation.  $h^2$ : heritability

**Materials and methods**

**Plant materials**

Two parents with large phenotypic differences in seed size and shape traits, *N24852* (*G. soja*) and *NN1138-2* (*G.*

*max*), were chosen to construct the CSSL population (Fig. 1a). The former, an annual wild soybean in maturity group (MG) III with a small seed size and inferior seed shape, was used as the donor parent, while the latter, an MG V with large seed size and superior seed shape, as well as high yield and oil content, was used as the

recurrent parent. A total of 1100 advanced backcrossed lines were developed by an integrated method of continuous backcross followed with alternation of backcross and selfing, including BC<sub>2</sub>F<sub>4</sub>, (BC<sub>2</sub>F<sub>3</sub>)BC<sub>1</sub>F<sub>2</sub>, (BC<sub>2</sub>F<sub>2</sub>)BC<sub>1</sub>F<sub>2</sub>, (BC<sub>2</sub>F<sub>2</sub>)BC<sub>2</sub>F<sub>2</sub>, BC<sub>3</sub>F<sub>3</sub>, BC<sub>4</sub>F<sub>2</sub> and BC<sub>5</sub>F<sub>2</sub> generations. Wang et al. (2013a) developed the first CSSL population *SojaCSSLP1* that was improved upon later by Xiang et al. (2015) and He et al. (2014) to form *SojaCSSLP2* and *SojaCSSLP3*, respectively, with an improved rate of *NN1138-2* genome recovery. The three continuously developed and improved CSSL populations were composed of 151, 153 and 158 lines with 151, 154 and 198 simple sequence repeat (SSR) markers, respectively. However, there were still many excessive wild and heterozygous segments unevenly distributed on the 20 chromosomes in these CSSL populations. The population was to be further improved based on these established ones and accumulated backcross materials. To reduce the percentage of redundant wild and heterozygous segments in the population, each *SojaCSSLP3* line were selfed and marker-assist-selected again. Furthermore, some lines reserved in the accumulated backcross materials were chosen as complementary sources to the new population. For wild segment identification, a single marker instead of double or multiple markers was used to represent wild segments (except a few lines with more than one marker) and to spread the markers more evenly on the 20 chromosomes. Finally, an improved CSSL population composed of 195 lines with 196 SSRs was constructed as *SojaCSSLP4*.

### Phenotype evaluation

The *SojaCSSLP4* population, *N24852* (*G. soja*) and *NN1138-2* (*G. max*), were planted in a complete randomized block design experiment with 3 replications, 3 rows per plot, a 0.5 m row width and 2.0 m row length, in 2011 and 2012 at Jiangpu Experimental Station of Nanjing Agricultural University, Nanjing, China (N31°02', E118°04') and Anhui Science and Technology University, Fengyang, China (N32°83', E117°55'). The experiments were coded as 2011JP, 2012JP, 2011FY and 2012FY. Field management included normal practices.

After the seeds were harvested and dried, the 5 seed size and shape traits, SL, SW, SR, SP and SA, were evaluated using the SC-G Grain Appearance Quality Image System developed by Hangzhou WSeen Detection Technology Co., Ltd, Hangzhou, China. SL was defined as the longest distance across the seed parallel to the hilum; SW as the longest distance across the seed perpendicular to the hilum; SP as the edge length of a seed; SA was the seed cross section area and SR was calculated from SW/SL (Fig. 1b).

### SSR marker analysis

The fresh leaves of the derived materials were collected, from which genomic DNA was extracted using the CTAB method (Doyle and Doyle 1990; Rogers and Bendich 1985) with minor modifications. From the high density integrated soybean genetic linkage map (Song et al. 2004), a total of 196 simple sequence repeat (SSR) markers that showed polymorphism between *N24852* (*G. soja*) and *NN1138-2* (*G. max*), were used to genotype the *SojaCSSLP4*.

Polymerase chain reaction (PCR) was conducted according to the method reported by Panaud et al. (1996) with minor modifications. The optimal system was in a 10 μL reaction volume containing about 1.5 μl of 10 × PCR buffer (0.1% gelatin; 500 mmol/L KCl; 100 mmol/L Tris-HCl, pH = 8.0), 1.2 μl of 2.5 mmol/L MgCl<sub>2</sub>, 0.24 μL of 10 mol/μL dNTP (N = A, C, G, T), 50 ng of genomic DNA, 0.12 μl of Taq polymerase (5 U/μl), 1.5 μl of each primer (0.2 mmol/L), with double-distilled water added to 10 μL. The PCR protocol was as follows: the template DNA of soybean was pre-denatured at 95 °C for 5 min, followed by 30 cycles of denaturation at 94 °C for 30 s, annealing at 55 °C for 30 s, extension at 72 °C for 40 s and a final extension at 72 °C for 10 min. Then, PCR products were cooled to 12 °C. The polymorphism of PCR product was detected on 8% PAGE gel. Gels were run in 1 × TBE (PAGE running buffer) for 1.2 h at 200 V and then silver-stained.

### Data analysis

Analysis of variance (ANOVA) of the 5 seed size and shape traits was performed using PROC GLM of SAS 9.1 (SAS Institute Inc., Cary, NC). The statistical model was  $Y = \mu + \alpha_i + \beta_j + \delta_l + (\alpha\beta)_{ij} + (\alpha\delta)_{il} + (\beta\delta)_{jl} + (-$   
 $\Psi\theta\epsilon\sigma\epsilon \alpha_i \text{ ιτ υθε εζζεγν οζ υθε ινθ μιξε β}_j \text{ ιτ υθε εζζεγν οζ υθε } j\upsilon\theta \text{ τινε } \delta_l \text{ ιτ υθε εζζεγν οζ υθε λθ εασ } \alpha\beta_{ij} \text{ ιτ υθε ηεξοσσ } \upsilon\pi\epsilon \times \text{ τινε } \zeta\upsilon\epsilon\sigma\alpha\gamma\upsilon\iota\omicron\zeta \alpha\delta_{il} \text{ ιτ υθε ηεξοσσπε } \times \text{ εασ } \zeta\upsilon\epsilon\sigma\sigma \alpha\gamma\upsilon\iota\omicron\zeta \beta\delta_{jl} \text{ ιτ υθε εασ } \times \text{ τινε } \zeta\upsilon\epsilon\sigma\alpha\gamma\upsilon\iota\omicron\zeta\kappa \alpha\beta\delta_{ijl} \text{ ιτ υθε } \eta\epsilon\zeta\upsilon\pi\epsilon \times \text{ τινε } \times \text{ εασ } \zeta\upsilon\epsilon\sigma\alpha\gamma\upsilon\iota\omicron\zeta \gamma_k \text{ ιτ υθε εζζεγν οζ κνθ } \sigma\epsilon\pi\mu\gamma\alpha\upsilon\iota\omicron\zeta \epsilon \text{ ιτ υθε } \sigma\epsilon\tau\iota\delta\phi\alpha\mu\phi \text{ Αμμ εζζεγντ } \psi\epsilon\sigma\epsilon \gamma\omicron\zeta\pi \text{ τιδεσεδ } \upsilon\theta \text{ βε } \sigma\alpha\zeta\delta\omicron\nu\phi \text{ Υθε } \theta\epsilon\sigma\omega\alpha\beta\iota\mu\upsilon \text{ h } \alpha\zeta\delta \text{ ηεξευνι } \gamma\omicron\epsilon\zeta\zeta\iota\gamma\iota\epsilon\zeta\upsilon \text{ οζ } \chi\alpha\sigma\iota\alpha\upsilon\iota\omicron\zeta \text{ GCV } \psi\epsilon\sigma\epsilon \text{ ετινινεδ } \alpha\tau \zeta\omicron\mu\mu\omicron\psi\tau \text{ h}^2 = \sigma_g\phi\sigma_g + \sigma_{gs}\phi\gamma + \sigma_{g\gamma}\phi\sigma + \sigma_{\phi\sigma\gamma\kappa} \text{ GCV} = \sigma_g/ \mu \times \zeta\phi \Psi\theta\epsilon\sigma\epsilon \text{ h}^2 \text{ ιτ υθε } \theta\epsilon\sigma\omega\alpha\beta\iota\mu\upsilon \text{ οζ υθε } \psi\theta\omicron\mu\epsilon \text{ εωπεσινεζυ } \sigma_g^2 \text{ ιτ υθε } \eta\epsilon\zeta\upsilon\pi\iota\gamma \text{ } \chi\alpha\sigma\iota\alpha\zeta\gamma\epsilon \text{ } \alpha\nu\omicron\zeta\eta \text{ ΣΙΜτ } \sigma_{gs} \text{ ιτ υθε } \eta\epsilon\zeta\upsilon\pi\epsilon \times \text{ τινε } \zeta\upsilon\epsilon\sigma\alpha\gamma\upsilon\iota\omicron\zeta \text{ } \chi\alpha\sigma\iota\alpha\zeta\gamma\epsilon \text{ } \sigma_{g\gamma} \text{ ιτ υθε } \eta\epsilon\zeta\upsilon\pi\epsilon \times \text{ εασ } \zeta\upsilon\epsilon\sigma\alpha\gamma\upsilon\iota\omicron\zeta \text{ } \chi\alpha\sigma\iota\alpha\zeta\gamma\epsilon \text{ } \sigma \text{ ιτ υθε } \epsilon\sigma\sigma\sigma \text{ } \chi\alpha\sigma\iota\alpha\zeta\gamma\epsilon \text{ } \mu \text{ ιτ υθε } \nu\epsilon\alpha\zeta \text{ οζ υθε } \rho\omicron\pi\phi\mu\alpha\upsilon\iota\omicron\zeta \text{ Χαι } \zeta\phi$

The genotypic graphics of the CSSLs were implemented by using the software Graphical Geno-Types2 (GGT2) (van Berloo 2008).

Following the mapping strategy of Su et al. (2010), SMA (single marker analysis) (SAS 2004), ICIM (inclusive composite interval mapping, RSTEP-LRT-ADD model) of IciMapping Version 2.0 (Li et al. 2007) and MCIM (mixed linear composite interval mapping, mixed linear model) of QTLNetwork version 2.0 (Yang et al. 2007) were used to detect the wild segments that confer the 5 seed size and shape traits. Thresholds of SMA for QTL detection were set as  $P \leq 0.05$  while that of IciMapping for QTL detection were also set as  $P \leq 0.05$  using 1000 permutations at 1-cM intervals.

## Results

### The genomic constitution of the improved CSSL population *SojaCSSLP4*

A total of 195 CSSLs were included in the *SojaCSSLP4* with 196 SSR markers, scattered throughout the 20 chromosomes, covering 2069.51 cM of soybean genome and an average distance of 11.92 cM between neighbouring markers (ranging from 0.38 cM to 41.37 cM) based on the published genetic linkage map by Song et al. (2004).

The genotypes of the substitution segments for each of the 195 CSSLs are shown in Fig. 2 using GGT2 software. As an example, the segregation of the PCR products of a SSR marker (Satt551) among the CSSLs along with their parents is shown in Fig. 3, which demonstrates that the wild segment located in the cultivated background. In *SojaCSSLP4*, the main diagonal wild segment length ranging between 2.22 and 64.10 cM with an average length of 21.11 cM; those less than 16 cM accounted for 48.47% of the diagonal segments (Fig. 2). This population was comprised of 892 wild segments and 376 heterozygous segments, with the wild segment length ranging from 2.22 to 69.50 cM in an average of 12.01 cM; the heterozygous segment length ranging from 2.22 to 71.80 cM in an average of 11.14 cM. The increased number of lines reduced the single segment size, which was the most significant improvement of the population.

Table 1 shows the genomic constitution of *SojaCSSLP4* in comparison with the 3 previous *SojaCSSL* populations. In the whole population, the recurrent parent genome recovery rate was 95.33% while the genomic portion of heterozygous segments was 0.98%, the wild genome segments portion was 3.67% and the genome missing rate was 0.02%. In an ideal SSSL population, the number of lines should be  $>300$ ; the theoretical recurrent parent genome recovering rate should be  $1-1/300$ ; the donor parent genome rate should be  $1/300$ ; the heterozygous rate and missing genome rate should be close to 0. In the new *SojaCSSLP4* population, all these genomic constitution

statistics were further improved in comparison to the 3 previous *SojaCSSL* populations. However, *SojaCSSLP4* had only 6 SSSLs, still far from a complete SSSL population, therefore, further tremendous effort is required.

### Genetic variation of the 5 seed size and shape traits in *SojaCSSLP4*

The ANOVA results of the 5 seed size and shape traits showed significant variations among the lines, with even the recurrent parent genome portion recovered to 95.33% (Table 2). Significant differences were observed between the cultivated and wild parents in SL, SW, SR, SP and SA (Fig. 1a). As shown in Fig. 1c–g, the mean values of the population were close to that of the recurrent parent *NN1138-2* for all 5 traits, which indicated that a majority of the CSSL genome had been recovered to the *NN1138-2* parent. In each environment, the coefficients of variation were less than 15% (data not shown; can be estimated from Table 2 and Fig. 1c–g), indicating that experimental precision was well-controlled. Meanwhile, broad ranges of the 5 traits demonstrated the wide variations among CSSLs, total of 7 (CSSL13, 44, 141, 142, 153, 164 and 188), 8 (CSSL13, 19, 44, 129, 141, 155, 164 and 171), 6 (CSSL19, 22, 129, 167, 171 and 187), 5 (CSSL13, 19, 44, 141 and 164) and 7 (CSSL13, 19, 44, 141, 142, 155 and 164) lines were significantly different from the recurrent parent *NN1138-2* in SL, SW, SR, SP and SA, respectively, under the 4 environments.

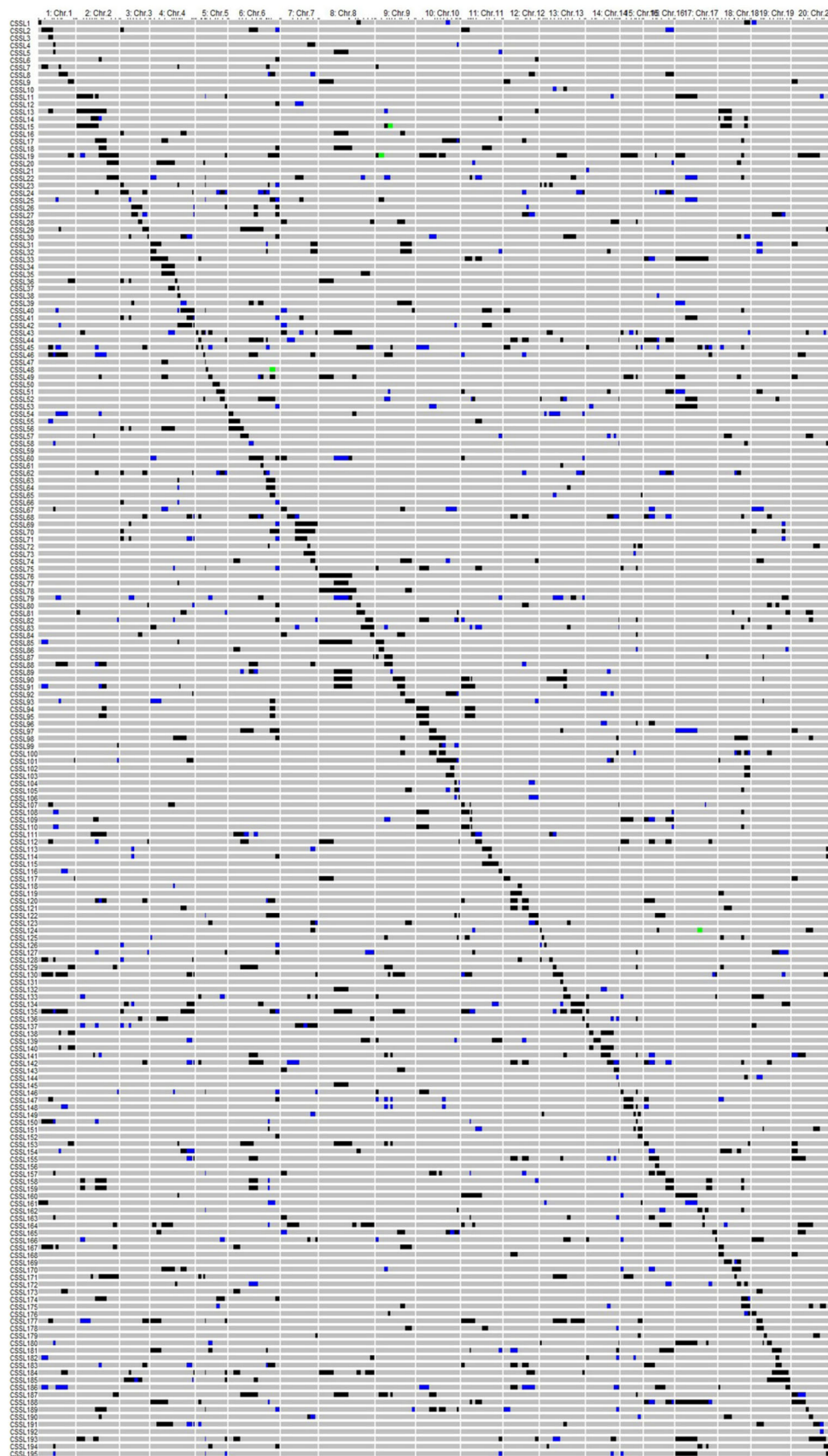
The heritability values of SL, SW, SR, SP, and SA were high, with averages of 83.48, 78.98, 80.31, 77.69 and 78.08% under the 4 environments (Fig. 1c–g), respectively. This indicated that the genetic variation accounted for a major part of the phenotypic variance in the CSSL population for the 5 traits.

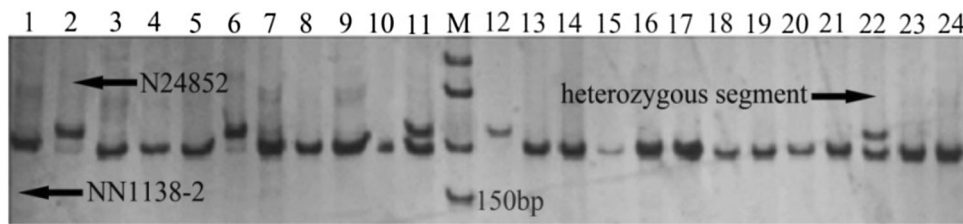
Table 3 indicates that in the population, SR showed a slight positive correlation with SW, but slight negative correlation with SL and no correlation with SP and SA. All correlations among SL, SW, SP and SA were positive, high and extremely significant in *SojaCSSLP4*. That means that any one of the 4 traits could be used as their representative in the improvement of the 4 traits; however, seed roundness is somewhat independent from the 4 traits.

### Wild segments that associate with seed size and shape traits in *SojaCSSLP4*

Significant variation among the lines for the 5 traits supported the segment mapping studies. Table 4 and Fig. 4 show the mapping results using SMA, ICIM and MCIM procedures in *SojaCSSLP4*. For SL, a total of 11 wild segments (Satt296, Sat\_140, **Satt305**, **Satt540**, Satt273, Sat\_270, Sat\_424, **Satt249**, Satt488, Sat\_223 and **Satt239**;

**Fig. 2** Graphical genotypes of the *SojaCSSLP4*. Each row represents a CSSLine and each column represents a chromosome (linkage group). The grey bar and black bars denote the segments homozygous for *NN138-2* and *N24852*, respectively; heterozygous segments were denoted with blue bar, and the missing ones were denoted with green bar





**Fig. 3** Amplification of a SSR marker (Satt551) in *SojaCSSLP4* derived from a cross between *NN1138-2* and *N24852*. 1: Female parent *NN1138-2*; 2: Male parent *N24852*; 3–24: CSSLs; M: 50 bp Marker

**Table 2** ANOVA of seed size and shape traits in *SojaCSSLP4*

Source of variation	DF	SL		SW		SR		SP		SA	
		MS	F value	MS	F value	MS	F value	MS	F value	MS	F value
CSSL	195	0.70	28.50**	0.54	17.12**	0.00	14.53**	6.18	14.79**	63.15	25.29**
Site	1	134.10	2261.40**	71.86	2568.70**	0.04	165.40**	1089.2	2507.20**	11,817.2	2731.70**
Year	1	45.80	772.60**	19.21	686.80**	0.04	152.40**	367.10	845.10**	3763.2	869.90**
CSSL × Site	195	0.07	1.11	0.03	1.15	0.00	1.44**	0.47	1.09	4.70	1.09
CSSL × Year	184	0.11	1.88**	0.06	1.64**	0.00	1.36**	0.87	1.39**	8.79	1.91**
CSSL × Year × Site	183	0.07	1.10	0.03	1.07	0.00	1.33**	0.47	1.09	4.85	1.12
Error	1482	0.06		0.02		0.00		0.44		4.34	

SL, SW, SR, SP and SA represent seed length, seed width, seed roundness, seed perimeter and seed cross section area, respectively; DF represents degree of freedom; MS represents mean of square

\* and \*\* represent significance at 0.05 and 0.01, respectively

**Table 3** Correlation coefficients among the five seed size and shape traits in *SojaCSSLP4*

Traits	SL	SW	SR	SP
SW	0.78**			
SR	-0.28**	0.37**		
SP	0.96**	0.92**	-0.21	
SA	0.95**	0.94**	0.03	1.00**

SL, SW, SR, SP and SA represent seed length, seed width, seed roundness, seed perimeter and seed cross section area, respectively

\*\* Represent the significantly correlation levels at 0.01

those in boldface had a contribution >9% in this section) were detected by using SMA, located on Chr. 2, Chr. 4, Chrs 6–7, Chr. 9, Chr. 11, Chr. 14, Chrs 16–18 and Chr. 20 (Fig. 4), respectively. Among them, 7 and 6 wild segments were also detected by ICIM and MCIM, respectively. On the basis of the SMA results, the phenotypic variation (PV) accounted for by individual QTL/segment ranged from 3.17 to 11.37% with some portions overlapped, and the alleles from wild soybean were all related to negatively additive effects (ADD) ranging from -0.86 to -0.18 mm (Table 4).

For SW, a total of 13 wild segments (Sat\_160, **Sat\_069**, Satt194, **Satt305**, **Satt540**, Satt055, Sat\_270, Satt663, Satt384, Satt249, Satt488, Sat\_223 and **Satt239**) were

identified using SMA, located on Chrs 1–2, Chr. 4, Chrs 6–7, Chr. 9, Chr. 11, Chr. 13, Chrs 15–18 and Chr. 20 (Fig. 4). Among them, 11 and 5 wild segments were also detected by ICIM and MCIM, respectively. On the basis of the SMA results, the PV accounted for by individual QTL/segment ranged from 3.54 to 20.26% with some portion overlapped and the alleles from wild soybean were all related to negative effects with ADDs ranging from -0.19 to -0.80 mm (Table 4).

For SR, a total of 7 wild segments of Sat\_160, **Satt274**, Staga001, Satt055, Sat\_282, Satt663 and Satt239 were identified by using SMA, distributed on Chr. 1, Chr. 2, Chr. 6, Chrs 9–10, Chr. 13, Chr. 15 and Chr. 20 (Fig. 4). Among them, 2 and 3 wild segments were also detected by ICIM and MCIM, respectively. On the basis of the SMA results, the PV accounted for by individual QTL/segment ranged from 4.39 to 56.48% with some portion overlapped and the alleles from wild soybean were all related to negative effects with ADDs ranging from -0.02 to -0.07 (Table 4). Here it should be noted that segment Satt274 on Chr. 2 made the largest contribution to phenotypic variation (56.48%) and was the most important segment for SR.

A total of 15 wild segments (Sat\_069, Sat\_239, Sat\_140, **Satt305**, **Satt540**, Satt508, Satt273, Sat\_270, Satt253, Satt663, Satt384, **Satt249**, Satt488, Sat\_223 and **Satt239**) were identified to have association with SP by using SMA,

**Table 4** The wild segments associated with seed size and shape traits in *SojaCSSLP4*

QTL	Segment	Position (cM)	<i>p</i>	PVE (%)	ADD	Method	Reference
<i>qSL2</i>	Satt296	52.61	0.001	6.07	−0.29	abc	Niu et al. (2013)
<i>qSL4</i>	Sat_140	41.43	0.007	3.67	−0.20	a	Salas et al. (2006)
<i>qSL6</i>	Satt305	69.67	<0.001	9.07	−0.27	abc	Salas et al. (2006)
<i>qSL7</i>	Satt540	35.85	<0.001	11.37	−0.86	ac	Salas et al. 2006
<i>qSL9</i>	Satt273	56.62	0.006	3.80	−0.36	ac	Niu et al. (2013)
<i>qSL11</i>	Sat_270	21.99	<0.001	8.00	−0.31	ab	
<i>qSL14</i>	Sat_424	100.10	0.010	3.45	−0.18	abc	
<i>qSL16</i>	Satt249	11.74	<0.001	9.26	−0.36	abc	Niu et al. (2013)
<i>qSL17</i>	Satt488	89.20	0.005	4.09	−0.37	a	Niu et al. (2013)
<i>qSL18</i>	Sat_223	61.64	0.013	3.17	−0.21	abc	
<i>qSL20</i>	Satt239	36.94	<0.001	9.03	−0.42	a	
<i>qSW1</i>	Sat_160	104.28	0.003	4.47	−0.26	ac	
<i>qSW2</i>	Sat_069	102.60	<0.001	20.26	−0.64	abc	Niu et al. (2013)
<i>qSW4</i>	Satt194	26.35	0.004	4.24	−0.20	a	
<i>qSW6</i>	Satt305	69.67	<0.001	11.10	−0.26	abc	Salas et al. (2006)
<i>qSW7</i>	Satt540	35.85	<0.001	13.33	−0.80	ac	Salas et al. (2006)
<i>qSW9</i>	Satt055	32.96	0.002	5.03	−0.27	ac	
<i>qSW11</i>	Sat_270	21.99	<0.001	7.23	−0.25	ab	
<i>qSW13</i>	Satt663	56.17	<0.001	7.89	−0.37	ac	
<i>qSW15</i>	Satt384	19.30	<0.001	7.45	−0.31	ac	
<i>qSW16</i>	Satt249	11.74	<0.001	8.44	−0.30	abc	
<i>qSW17</i>	Satt488	89.20	0.002	4.87	−0.35	ac	Niu et al. (2013)
<i>qSW18</i>	Sat_223	61.64	0.008	3.54	−0.19	ac	
<i>qSW20</i>	Satt239	36.94	<0.001	18.17	−0.51	abc	Niu et al. (2013)
<i>qSR1</i>	Sat_160	104.28	<0.001	7.44	−0.03	ab	
<i>qSR2</i>	Satt274	116.35	<0.001	56.48	−0.07	abc	Niu et al. (2013)
<i>qSR6</i>	Staga001	119.85	<0.001	6.63	−0.02	abc	
<i>qSR9</i>	Satt055	32.96	<0.001	7.78	−0.03	a	Niu et al. (2013)
<i>qSR10</i>	Sat_282	63.81	<0.001	6.76	−0.04	a	
<i>qSR13</i>	Satt663	56.17	0.002	4.69	−0.02	a	
<i>qSR20</i>	Satt239	36.94	0.003	4.39	−0.02	a	
<i>qSP2</i>	Sat_069	102.60	0.001	5.74	−1.17	a	
<i>qSP3</i>	Sat_239	87.35	0.012	3.21	−0.80	a	
<i>qSP4</i>	Sat_140	41.43	0.003	4.40	−0.65	ac	
<i>qSP6</i>	Satt305	69.67	<0.001	10.72	−0.88	abc	
<i>qSP7</i>	Satt540	35.85	<0.001	13.68	−2.79	ac	
<i>qSP8</i>	Satt508	108.78	0.009	3.49	−1.02	a	
<i>qSP9</i>	Satt273	56.62	0.009	3.48	−1.02	a	
<i>qSP11</i>	Sat_270	21.99	<0.001	8.58	−0.94	a	
<i>qSP12</i>	Satt253	67.17	0.002	5.24	−0.76	a	
<i>qSP13</i>	Satt663	56.17	0.004	4.30	−0.93	ac	
<i>qSP15</i>	Satt384	19.30	0.001	5.75	−0.94	a	
<i>qSP16</i>	Satt249	11.74	<0.001	9.75	−1.09	abc	
<i>qSP17</i>	Satt488	89.20	0.002	5.03	−1.23	ac	
<i>qSP18</i>	Sat_223	61.64	0.006	3.87	−0.69	ac	
<i>qSP20</i>	Satt239	36.94	<0.001	13.83	−1.54	ab	
<i>qSA2</i>	Sat_069	102.60	<0.001	7.06	−4.15	ac	
<i>qSA4</i>	Sat_140	41.43	0.005	4.09	−2.00	ac	
<i>qSA6</i>	Satt305	69.67	<0.001	11.11	−2.84	abc	



**Table 4** continued

QTL	Segment	Position (cM)	<i>p</i>	PVE (%)	ADD	Method	Reference
<i>qSA7</i>	Satt540	35.85	<0.001	13.04	−8.69	ac	
<i>qSA8</i>	Satt508	108.78	0.014	3.09	−3.07	a	
<i>qSA9</i>	Satt273	56.62	0.010	3.41	−3.22	ac	
<i>qSA11</i>	Sat_270	21.99	<0.001	8.41	−2.98	ac	
<i>qSA12</i>	Satt253	67.17	0.001	5.28	−2.43	a	
<i>qSA13</i>	Satt663	56.17	0.002	4.89	−3.16	ac	
<i>qSA15</i>	Satt384	19.30	0.001	6.05	−3.06	ac	
<i>qSA16</i>	Satt249	11.74	<0.001	9.65	−3.47	c	
<i>qSA17</i>	Satt488	89.20	0.003	4.52	−3.71	c	
<i>qSA18</i>	Sat_223	61.64	0.007	3.68	−2.15	c	
<i>qSA20</i>	Satt239	36.94	<0.001	13.85	−4.92	b	

PVE represents percentage of phenotypic variation explained by the QTL; ADD represents additive effect of the allele from *G. soja*

SL, SW, SR, SP and SA represent seed length, seed width, seed roundness, seed perimeter and seed cross section area, respectively

In the column of Method, “a” represents the QTL/segment was detected using SMA procedure; “b” represents the QTL/segment was detected using MCIM procedure; “c” means the QTL/segment was detected using ICIM procedure

In the column of Reference, the literature of the reported QTL segment was given

located on Chrs 2–4, Chrs 6–9, Chrs 11–13, Chrs 15–18 and Chr. 20 (Fig. 4). Among them, 7 and 3 wild segments were also detected by ICIM and MCIM, respectively. On the basis of the SMA results, the PV accounted for by individual QTL/segment ranged from 3.21 to 13.83% with some portion overlapped and the alleles from wild soybean were all related to negative effects with ADDs from −2.79 to −0.65 mm (Table 4).

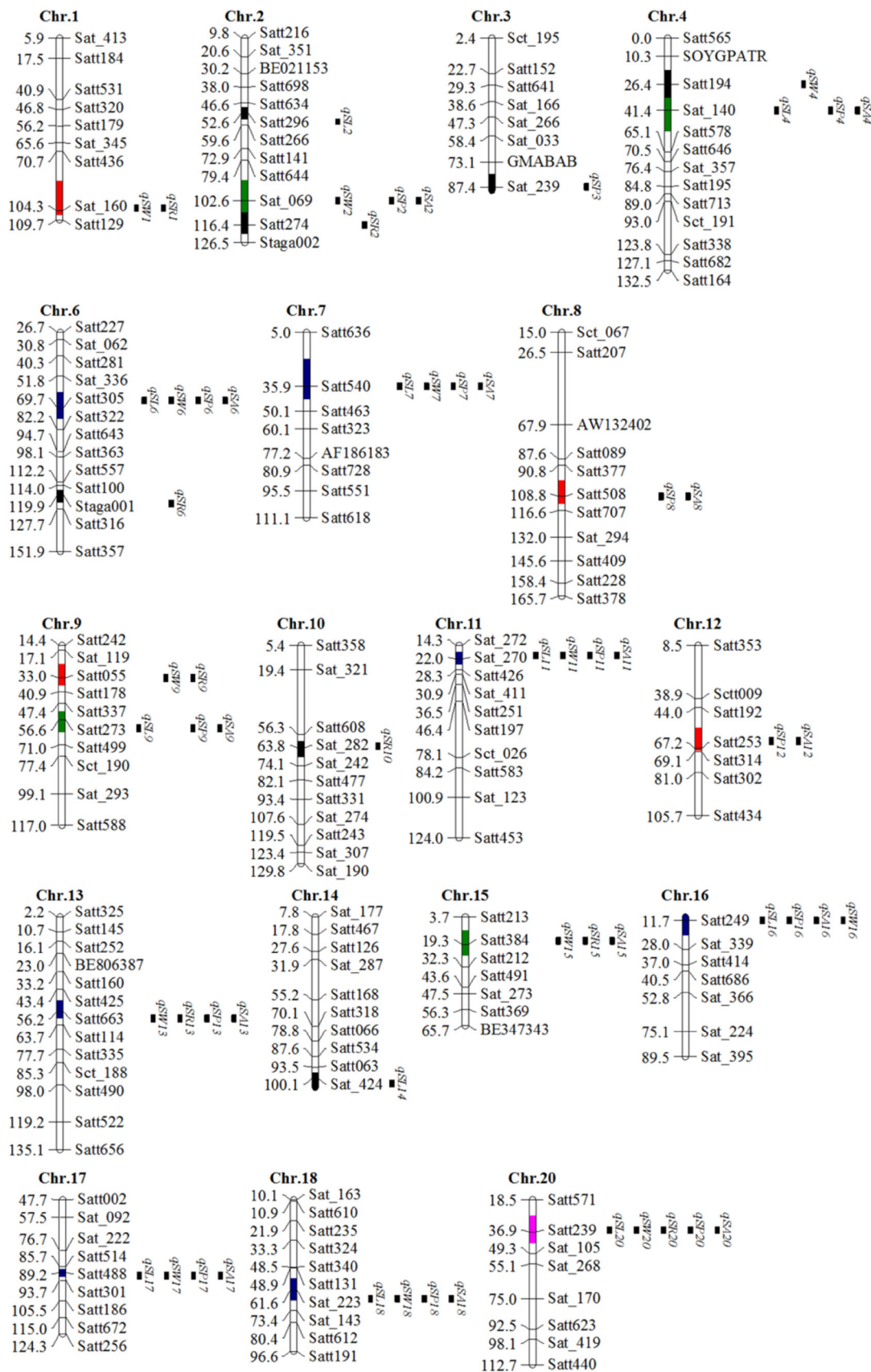
A total of 14 wild segments (Sat\_069, Sat\_140, **Satt305**, **Satt540**, Satt508, Satt273, Sat\_270, Satt253, Satt663, Satt384, **Satt249**, Satt488, Sat\_223 and **Satt239**) were identified to have association with SA by using SMA, located on Chr. 2, Chr. 4, Chrs 6–9, Chrs 11–13, Chrs 15–18 and Chr. 20 (Fig. 4). Among them, 12 and 2 wild segments were also detected by ICIM and MCIM, respectively. On the basis of SMA results, the PV accounted for by individual QTL/segment ranged from 3.09 to 13.85% with some portion overlapped and the alleles from wild soybean were all associated with negative effects with ADD from −2.00 to −4.92 mm<sup>2</sup> (Table 4).

### Wild segments conferring multiple traits

According to the significant correlations among the 5 traits, a corresponding genetic basis was assumed to underlie the correlations (Table 3). From the mapping results in *SojaCSSLP4*, the wild segments that associated with multiple traits were detected in the CSSL population. As shown in Fig. 4, a total of 16 wild segments were

associated with at least 2 traits among the 60 detected wild segments distributed on all the chromosomes except Chr. 5 and Chr. 19. The major shared segments were as follows: (1) The wild segment of Satt239 on Chr. 20 conferred all 5 traits; (2) the wild segments of Satt305, Satt540, Sat\_270, Satt249, Satt488 and Sat\_223 were associated with 4 traits (SL, SW, SP and SA) and the wild segment of Satt663 was associated with SW, SR, SP and SA; (3) the wild segments of Sat\_069, Sat\_140, Satt273 and Satt384 were associated with 3 traits; and (4) Sat\_160, Satt508, Satt055 and Satt253 were associated with 2 traits. The above shared QTL/wild-segments, such as Satt239 on Chr. 20, Satt305 on Chr. 6, Satt540 on Chr. 7, Sat\_270 on Chr. 11, Satt663 on Chr. 13, Satt249 on Chr. 16, Satt488 on Chr. 17, and Sat\_223 on Chr. 18, should be the important genetic factors that control soybean seed size and shape. The finding of one wild segment associated with multiple traits can be called segment pleiotropism, which might be due to the different genes being closely located in the same region or might be due to some gene(s) with pleiotropism.

Among the detected QTLs/segments for SL, SW, SP, SA and SR, 2 of 11, 1 of 13, 1 of 15, 0 of 14 and 3 of 7 were independent, respectively, the others were shared with at least one or more traits. Here SR had three independent QTLs (Satt274, Staga001, Sat\_282) among the total seven loci/segments, more than that of the other traits, and the shared QTL/segments were mainly with seed width. This offers a genetic explanation for the above correlations.



◀**Fig. 4** Mapped QTL/segments of soybean seed size and shape traits on chromosomes in *SojaCSSLP4*. The linkage map follows that of Song et al. (2004). (black bar), (red bar), (green bar), (blue bar) and (pink bar) colours on a chromosome represent segment related to one, two, three, four and five seed size and shape traits, respectively

## Discussion

### Features of *SojaCSSLP4* and its implication to the utilization of wild resources

The recovery rate of the *NN138-2* genome was between 81.90 and 99.80% among the 195 lines in *SojaCSSLP4*, with an average of 95.33%. In comparison with the previous 3 *SojaCSSL* populations, the recurrent genome rate in *SojaCSSLP4* was raised with the wild genome rate reduced at a certain (but not large) amount (Table 1). The wild alleles/segments on the 196 markers existed in at least one CSSL and the 100% wild soybean genome substitution took place in *SojaCSSLP4*.

Potential applications of CSSLs covering an entire wild species genome was thoroughly discussed in Tanksley and Nelson (1996). Eshed and Zamir (1994) pointed out that using a wild relative as a donor parent to construct CSSLs is a more powerful application than using cultivated donors. First, CSSLs could be utilized as a permanent mapping population to readily map QTLs of quantitative traits, as in the present study, 11, 13, 7, 15 and 14 QTLs/segments linked to SL, SW, SR, SP and SA were identified, respectively. Second, CSSLs can be utilized for QTL fine-mapping and map-based gene cloning (Wan et al. 2008). According to the primary mapping results, QTL(s) could be fine-mapped using a residual heterozygous lines (RHL) population or an  $F_2$  and/or  $F_{2:3}$  population derived from a cross between the CSSL with a target QTL and the recurrent parent. In this way, the target segment is decomposed into one or several Mendelian factor(s), therefore the QTL(s) can be mapped to a small region(s). For example, in our previous study (He et al. 2015), *L1* gene for black pod colour was fine-mapped using a RHL population developed from the *SojaCSSLP3* (an early version of the *SojaCSSLP4*). The locus *L1* was narrowed down from a region of 52.48 to 7.8 cM, then further to 3.1 cM and finally to a 184.43-kb region on Chr.19 where located 13 putative genes. Among them, Glyma19g27460 was recognized, the possible candidate for the *L1* gene through comparing the qRT-PCR expression levels of the 13 candidate genes. Third, using the QTL/segment information, beneficial donor QTLs/segments can be integrated and transgressive genotypes can be designed. Therefore, *SojaCSSLP4* can provide novel germplasm in molecular breeding for elite seed traits through pyramiding QTLs/segments.

### Genetic differentiation between cultivated and wild soybeans in seed size and shape traits

During the domestication process, seed size and shape changed greatly from the wild to the cultivated soybean (Zhou et al. 2015). The wild soybean chromosome segment substitution line population is powerful for detecting allele differences between wild and cultivated soybeans in these traits. According to the present mapping results, a complex genetic system of seed size and shape traits in wild soybean were detected, and large differences in the 5 traits between cultivated and wild soybeans were found to be due to 11, 13, 7, 15 and 14 wild segments, respectively, with large or small phenotypic contribution (from 3.09 to 56.48%). The wild alleles of seed size and seed shape on the detected loci were all associated with negative or inferior effects, which were consistent with the phenotype difference between the wild and cultivated parents. Previous authors also used the wild soybean to detect QTL for seed size-related traits, such as 100 seed weight (Concibido et al. 2003). They found that all the detected wild alleles were associated with negative effects, which indicated that the selection of seed size might be unidirectional and that it became positive during the domestication process. Accordingly, the wild alleles of seed size and shape traits should be carefully selected against in introgression of *G. soja* into cultivated soybeans. Meanwhile, multiple traits sharing a same segment or a segment with pleiotropic effects were found on 16 wild segments. This was consistent with the correlation results among the 5 seed size and shape traits.

### Comparison of the detected QTL/segments with those reported in the literature

It is commonly understood that when mapping procedure is used to detect some QTLs for the CSSL population, the QTL is in fact mapped on some segment. Since the segment is a piece of chromosome, there might be a number of QTLs located on the same segment. With the help of Williams 82 Physical Maps (Schmutz et al. 2010), the QTLs detected in different linkage maps could be compared. In the literature, a total of 14 QTLs were detected in other crosses derived from cultivated soybeans (Table 4). Of the 11 QTLs/segments for SL, *qSL2*, *qSL9*, *qSL16* and *qSL17* were near or in the same regions as the QTLs detected in cultivated soybeans through association mapping (Niu et al. 2013), and *qSL4*, *qSL6* and *qSL7* were near or in the same regions as the QTLs from the cultivated soybean RIL population derived from Noir1 × Archer and Minsoy × Noir1 (Salas et al. 2006). Of the 13 QTLs/segments for SW, *qSW2* was also detected in cultivated soybeans through genome-wide association analysis (Niu et al. 2013), *qSW6* and *qSW7* were detected in the cultivated

RILs of Minsoy × Noir1 (Salas et al. 2006), *qSW17* were detected in the RILs of Noir1 × Archer (Salas et al. 2006) and *qSW20* were detected in cultivated soybeans through association mapping (Niu et al. 2013). Of the 8 QTLs/segments for SR, *qSR2* and *qSR9* were also detected in cultivated soybeans through association mapping (Niu et al. 2013). The above results indicated that allele differentiation happened not only between wild and cultivated soybeans, but also between cultivated parents on these seed size and shape QTLs. Accordingly, these QTLs/segments might be the most important loci with further differentiation taking place after domestication. Meanwhile, a number of QTLs were newly detected in the wild soybean *N24852*, including 4 for SL, 8 for SW and 5 for SR. It means that these QTLs/segments differentiated during domestication of the wild soybean and were historically old events which should be quite different from those detected between cultivated soybeans and should be studied further.

The phenomenon of a segment conferring multiple traits has been reported previously in several crops, including rice, maize, soybean, and so on (Bian et al. 2010; Hu et al. 2013; Liu et al. 2014). In this study, the QTLs *qSL-20*, *qSW-20*, *qSR-20*, *qSP-20* and *qSA-20* for SL, SW, SR, SP and SA were mapped to a common genomic region of Chr. 20. This segment pleiotropism needs to be studied further.

In addition, from the present results, it seems that two different but related sets of segments or genetic systems for seed size and shape traits exist in the genome, one for SL, SW, SP and SA (containing major shared segments Satt305, Satt540 and Satt239), another for SR (containing the major independent segment Satt274). Therefore, in the improvement of seed size and shape in utilization of annual wild soybean, the 2 genetic systems of seed size and shape traits should be controlled simultaneously and Satt305 on Chr. 6, Satt540 on Chr. 7, Satt239 on Chr.20 and Satt274 on Chr. 2 are the most important segments for further study.

In the above text, we used QTL/segments to represent the mapping target. In fact, each segment of the CSSL population contains a large piece of chromosome or many QTL/genes. Thus, the identified wild segments may be understood as a macro-segment mapping result that is to be further dissected into smaller segments as well as corresponding QTLs/genes. Therefore, *SojaCSSLP4* has laid a material base for the continuing dissection of the wild genome.

**Acknowledgements** This work was funded by The National High Technology Research and Development Program of China (2012AA101106), the Natural Science Foundation of China (31271750), the MOE 111 Project (B08025), the MOE Program for Changjiang Scholars and Innovative Research Team in University (PCSIRT13073), the MOA Public Profit Program (201203026-4), the

MOA CARS-04 program, the Jiangsu Higher Education PAPD Program and the Jiangsu JCIC-MCP program.

#### Compliance with ethical standards

**Conflict of interest** All the authors declare that they do not have any conflict of interest with this work.

#### References

- Alpert KB, Tanksley SD (1996) High-resolution mapping and isolation of a yeast artificial chromosome contig containing *fw2.2*: a major fruit weight quantitative trait locus in tomato. *Proc Natl Acad Sci USA* 93:15503–15507
- Basra AS (1995) Seed quality: basic mechanisms and agricultural implications. Food Products Press, Binghamton
- Belknap JK (2003) Chromosome substitution strains: some quantitative considerations for genome scans and fine mapping. *Mamm Genome* 14:723–732
- Bernacchi D, Beck-Bunn T, Eshed Y, Lopez J, Petiard V, Uhlig J, Zamir D, Tanksley S (1998) Advanced backcross QTL analysis in tomato. I. Identification of QTLs for traits of agronomic importance from *Lycopersicon hirsutum*. *Theor Appl Genet* 97:381–397
- Bian JM, Jiang L, Liu LL, Wei XJ, Xiao YH, Zhang LJ, Zhao ZG, Zhai HQ, Wan JM (2010) Construction of a new set of rice chromosome segment substitution lines and identification of grain weight and related traits QTLs. *Breed Sci* 60:305–313
- Concibido VC, La Vallee B, Mclaird P, Pineda N, Meyer J, Hummel L, Yang J, Wu K, Delannay X (2003) Introgression of a quantitative trait locus for yield from *Glycine soja* into commercial soybean cultivars. *Theor Appl Genet* 106:575–582
- Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus* 12:13–14
- Eshed Y, Zamir D (1994) A genomic library of *Lycopersicon pennellii* in *L. esculentum*: a tool for fine mapping of genes. *Euphytica* 79:175–179
- Eshed Y, Zamir D (1995) An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics* 141:1147–1162
- Fatokun CA, Menancio-Hautea DI, Danesh D, Young ND (1992) Evidence for orthologous seed weight genes in cowpea and mung bean based on RFLP mapping. *Genetics* 132:841–846
- Frery A, Nesbitt TC, Frery A, Grandillo S, Van Der Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB (2000) *fw2.2*: a quantitative trait locus key to the evolution of tomato fruit size. *Science* 289:85–88
- Gai JY (2013) Experimental statistics. China Agriculture Press, Beijing
- He QY, Yang HY, Xiang SH, Wang WB, Xing GN, Zhao TJ, Gai JY (2014) QTL mapping for the number of branches and pods using wild chromosome segment substitution lines in soybean [*Glycine max* (L.) Merr.]. *Plant Genet Resour C* 12:S172–S177
- He QY, Yang HY, Xiang SH, Tian D, Wang WB, Zhao TJ, Gai JY (2015) Fine mapping of the genetic locus *L1* conferring black pods using a chromosome segment substitution line population of soybean. *Plant Breed* 134:437–445
- Hu ZB, Zhang HR, Kan GZ, Ma DY, Zhang D, Shi GX, Hong DL, Zhang GZ, Yu DY (2013) Determination of the genetic architecture of seed size and shape via linkage and association analysis in soybean (*Glycine max* L. Merr.). *Genetica* 141:247–254

- Hymowitz T (1990) Soybeans: the success story. *Advances in new crops*. Timber Press, Portland, pp 159–163
- Li HH, Ye GY, Wang JK (2007) A modified algorithm for the improvement of composite interval mapping. *Genetics* 175:361–374
- Liu J, Van Eck J, Cong B, Tanksley SD (2002) A new class of regulatory genes underlying the cause of pear-shaped tomato fruit. *Proc Natl Acad Sci USA* 99:13302–13306
- Liu Y, Wang LW, Sun CL, Zhang ZX, Zheng YL, Qiu FZ (2014) Genetic analysis and major QTL detection for maize kernel size and weight in multi-environments. *Theor Appl Genet* 127:1019–1037
- Mao H, Sun S, Yao J, Wang C, Yu S, Xu C, Li X, Zhang Q (2010) Linking differential domain functions of the *GS3* protein to natural variation of grain size in rice. *Proc Natl Acad Sci USA* 107:19579–19584
- Martin A, Lee J, Kichey T, Gerentes D, Zivy M, Tatout C, Dubois F, Balliau T, Valot B, Davanture M, Terce-Laforgue T, Quillere I, Coque M, Gallais A, Gonzalez-Moro MB, Bethencourt L, Habash DZ, Lea PJ, Charcosset A, Perez P, Murigneux A, Sakakibara H, Edwards KJ, Hirel B (2006) Two cytosolic glutamine synthetase isoforms of maize are specifically involved in the control of grain production. *Plant Cell* 18:3252–3274
- Moongkanna J, Nakasathien S, Novitzky W, Kwanyuen P, Sinchaisri P, Srinives P (2011) SSR markers linking to seed traits and total oil content in soybean. *Thai J Agric Sci* 44:233–241
- Niu Y, Xu Y, Liu XF, Yang SX, Wei SP, Xie FT, Zhang YM (2013) Association mapping for seed size and shape traits in soybean cultivars. *Mol Breed* 31:785–794
- Panaud O, Chen X, McCouch SR (1996) Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). *Mol Gen Genet* 252:597–607
- Rogers SO, Bendich AJ (1985) Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. *Plant Mol Biol* 5:69–76
- Salas P, Oyarzo-Llaipen JC, Wang D, Chase K, Mansur L (2006) Genetic mapping of seed shape in three populations of recombinant inbred lines of soybean (*Glycine max* L. Merr.). *Theor Appl Genet* 113:1459–1466
- SAS II (2004) SAS/STATs 9.1 User's Guide. SAS Publishing, Cary, NC, USA
- Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J (2010) Genome sequence of the palaeopolyploid soybean. *Nature* 463:178–183
- 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 XJ, Huang W, Shi M, Zhu MZ, Lin HX (2007) A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat Genet* 39:623–630
- Su CF, Zhao TJ, Gai JY (2010) Simulation comparisons of different effectiveness among QTL mapping procedures of different statistical genetic models. *Acta Agron Sin* 36(7):1100–1107
- Tanksley SD, Nelson JC (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theor Appl Genet* 92:191–203
- van Berloo R (2008) GGT 2.0: versatile software for visualization and analysis of genetic data. *J Hered* 99:232–236
- Van der Knaap E, Tanksley SD (2001) Identification and characterization of a novel locus controlling early fruit development in tomato. *Theor Appl Genet* 103:353–358
- Wan XY, Weng JF, Zhai HQ, Wang JK, Lei CL, Liu XL, Guo T, Jiang L, Su N, Wan JM (2008) Quantitative trait loci (QTL) analysis for rice grain width and fine mapping of an identified QTL allele *gw-5* in a recombination hotspot region on chromosome 5. *Genetics* 179:2239–2252
- Wang JK, Wan XY, Crossa J, Crouch J, Weng JF, Zhai HQ, Wan JM (2006) QTL mapping of grain length in rice (*Oryza sativa* L.) using chromosome segment substitution lines. *Genet Res* 88:93–104
- Wang WB, He QY, Yang HY, Xiang SH, Xing GN, Zhao TJ, Gai JY (2012) Detection of wild segments associated with number of branches on main stem and leafstalk angle in soybean. *Sci Agric Sin* 45:4749–4758
- Wang WB, He QY, Yang HY, Xiang SH, Zhao TJ, Gai JY (2013a) Development of a chromosome segment substitution line population with wild soybean (*Glycine soja* Sieb. et Zucc.) as donor parent. *Euphytica* 189:293–307
- Wang WB, He QY, Yang HY, Xiang SH, Zhao TJ, Xing GN, Gai JY (2013b) Identification of wild segments associated with stem termination, pod color, and seed coat color in soybean. *Acta Agron Sin* 39:1155–1163
- Xiang SH, Wang WB, He QY, Yang HY, Liu C, Xing GN, Zhan TJ, Gai JY (2015) Identification of QTL/segments related to agronomic traits using CSSL population under multiple environments. *Sci Agric Sin* 48:10–22
- Yang J, Zhu J, Williams RW (2007) Mapping the genetic architecture of complex traits in experimental populations. *Bioinformatics* 23:1527–1536
- Yang K, Moon JK, Jeong N, Chun HK, Kang ST, Back K, Jeong SC (2011) Novel major quantitative trait loci regulating the content of isoflavone in soybean seeds. *Genes Genom* 33:685–692
- Zamir E, Geiger B (2001) Molecular complexity and dynamics of cell-matrix adhesions. *J Cell Sci* 114:3583–3590
- Zhou ZK, Jiang Y, Wang Z, Gou ZH, Lyu J, Li WY, Yu YJ, Shu LP, Zhao YJ, Ma YM, Fang C, Shen YT, Liu TF, Li CC, Li Q, Wu M, Wang M, Wu YS, Dong Y, Wan WT, Wang X, Ding ZL, Gao YD, Xiang H, Zhu BG, Lee SH, Wang W, Tian ZX (2015) Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat Biotechnol* 33:408–414