Theor Appl Genet (2011) 122:865–874
DOI 10.1007/s00122-010-1492-5

ORIGINAL PAPER

Fine genetic mapping of the genomic region controlling leaflet shape and number of seeds per pod in the soybean

Namhee Jeong · Jung-Kyung Moon · Hong Sig Kim · Chang-Gi Kim · Soon-Chun Jeong

Received: 15 March 2010 / Accepted: 4 November 2010 / Published online: 23 November 2010

Abstract Narrow leaflet cultivars tend to have more seeds per pod than broad leaflet cultivars in soybean [*Glycine max* (L.) Merr.], which suggests that the leaflet-shape trait locus is tightly linked to or cosegregates with the trait locus controlling the number of seeds per pod (NSPP). Here, we attempted to further elucidate the relationship between leaflet shape and NSPP. A BC_3F_2 population from a cross between the 'Sowon' (narrow leaflets and high NSPP) and 'V94-5152' (broad leaflets and low NSPP) variants was used. The results of the molecular genetic analyses indicated that, although the NSPP characteristic, in particular, the occurrence of 4-seeded pods, is governed by additional modifying genes that are likely present in Sowon, the two traits cosegregate in the BC_3F_2 population. The mapping results generated using public markers demonstrated that the narrow leaflet-determining gene in Sowon is an allele of the previously highly studied *ln* gene on chromosome 20. A high-resolution map delimited the genomic region controlling both the leaflet shape and NSPP traits to a sequence length of 66 kb, corresponding to 0.7 cM. Among the three genes annotated in this 66 kb region, Glyma20g25000.1 appeared to be a good candidate for the *Ln*-encoding gene, owing to its 47.8% homology with the protein encoding for the *JAGGED* gene that regulates lateral organ development in *Arabidopsis*. Taken together, our results suggested that phenotypic variations for narrow leaflet and NSPP are predominantly from the pleiotropic effects of the *ln* gene. Thus, our results should provide a molecular framework for soybean breeding programs with the objective of improving soybean yield.

Communicated by H. T. Nguyen.

N. Jeong · C.-G. Kim · S.-C. Jeong (✉)
Bio-Evaluation Center, Korea Research Institute of Bioscience and Biotechnology, Cheongwon, Chungbuk 363-883, Republic of Korea
e-mail: scjeong@kribb.re.kr

J.-K. Moon
National Institute of Crop Science, Rural Development Administration, Suwon 441-857, Republic of Korea

H. S. Kim
Department of Crop Science, Chungbuk National University, Cheongju 361-763, Republic of Korea

Introduction

The selection of stable genotypes with increased seed yield is one of the most important goals of different breeding programs. Seed yield is determined by the number of seeds per unit area and seed weight. In the case of the soybean, the number of seeds per unit area is a product of the number of plants per unit area, the number of pods per plant, and the number of seeds per pod. Among these soybean yield components, the number of seeds per pod has long been suggested to be tightly associated with the soybean *ln* gene, which controls leaflet shape (Takahashi 1934; Domingo 1945).

Domingo (1945) studied crosses between broad (ovate) and narrow (lanceolate) leaflet types and suggested a crossover percentage of approximately 7.9 between the narrow leaflet gene and a gene presumed to control the number of seeds per pod (NSPP). On the other hand, Johnson and Bernard (1962) identified a high NSPP and narrow leaflet as pleiotropic effects of the same allele. Weiss (1970) concluded that the narrow leaflet allele exerts a major pleiotropic effect in increasing the frequency of 4-seeded pods, but the expression of the 4-seeded trait was decidedly affected

by the vigor of the plant and its modifying genes. Bernard and Weiss (1973) attributed the control of soybean leaflet shape to a single gene (*ln*), with broad dominant over narrow (Sawada 1988).

Due to an association between NSPP and leaflet shape, a number of studies have been performed to evaluate the effects of the narrow leaflet gene on seed yield. Mandl and Buss (1981) reported that broad and narrow leaflet soybean isolines showed the same seed yield, but narrow leaflet plants consistently showed smaller seeds than broad leaflet plants. Thus, it has been implied that the narrow leaflet plants produced a greater number of seeds, because they possessed more seeds per pod. You et al. (1995) also reported that, in an examination of 72 soybean cultivars to determine the effects of leaflet shape on seed yield and its components, no significant difference was observed between the broad and narrow leaflet lines in terms of average seed yield or number of pods per plant. However, significant differences for seed size and number of seeds per pod were noted. Dinkins et al. (2002) reported that, when yields and yield components within broad (*Ln/Ln*), heterozygous (*Ln/ln*), and narrow (*ln/ln*) leaflet types were compared in a population derived from crosses between broad and narrow leaflet soybean cultivars, the heterozygous leaflet plants generated significantly more pods per plant and higher plant yields in general than plants with either broad or narrow leaflets; they concluded that the heterozygous condition at the locus for leaflet morphology resulted in heterosis for plant yield. Although soybean yield has generally been described in terms of the total soybean seed weight per unit area (Pedersen and Lauer 2004), the result that the narrow leaflet plants tend to produce a greater number of seeds is particularly interesting because the number of seeds is a critical yield component for specialty soybean markets. For example, quantitative trait loci for soybean sprout yield are associated negatively with seed-weight loci (Lee et al. 2001), supporting the observation that small-seeded soybean cultivars tend to have higher soybean sprout yields (Kwon et al. 1981).

Collectively, narrow leaflet cultivars tend to have more seeds per pod than broad leaflet cultivars. The causes of this association might include either the tight linkage of genes that control independent traits or pleiotropic effects of the target gene for both of the traits. The principal objective of this study was to construct a high-resolution map of a chromosomal region that controls the narrow/broad leaflet trait, in order to elucidate in detail the genetic relationship between leaflet shape and NSPP. The results of molecular genetic analyses of NSPP, one of the components of soybean yield, should facilitate the development of novel soybean cultivars with high yield potential.

Materials and methods

Plant materials

A BC_3F_2 population from a cross between the soybean cultivars 'Sowon' and 'V94-5152' served as a segregating population for the establishment of linkage relationships between molecular markers and the narrow leaflet-shape trait presumed to be determined by the *ln* gene (hereafter, referred to as the SV population). Sowon was released for soybean sprouts in 1999, has a narrow leaflet shape, and generates approximately 70% 3-seeded pods and 5–10% 4-seeded pods (Park et al. 2000). V94-5152 was registered as a resistant germplasm that contains the *Rsv*4 allele conferring resistance to a broad range of soybean mosaic virus strains (Buss et al. 1997). V94-5152 has a broad leaflet shape and generates approximately 80% 2-seeded pods and 0% 4-seeded pods.

The SV population was developed in an effort to transfer the dominant *Rsv*4 allele from the donor parent, V94-5152 (*Rsv*4), to the recurrent parent, Sowon (*rsv*4). Sowon as females was crossed with V94-5152 as males to produce F_1 seeds in summer 2002 in the field at Suwon, Korea. Confirmed F_1 plants were backcrossed to Sowon as females to produce BC_1F_1 seeds in a greenhouse at Suwon during the winter of 2002/2003 and then the backcrossing of the BC_1F_1 to Sowon was repeated to produce BC_2F_1 seeds. During backcrossing, we made sure that, in addition to the *Rsv*4 gene, the narrow leaflet trait (the *ln* gene) of Sowon, which has been shown to be tightly linked with the increased NSPP trait, was retained. All F_1 and BC_1F_1 plants were screened with molecular markers (including Satt558 and Satt634) linked to *Rsv*4 (Hwang et al. 2006) and a few markers randomly chosen from the 20 soybean linkage groups (Song et al. 2004) in order to confirm that they are true crosses containing the *Rsv*4 gene. Their leaflet shape trait was assessed by visual inspection. All BC_2F_1 plants and their progeny were screened with the molecular markers linked to *Rsv*4 and a total of 80 molecular markers randomly chosen from each of the 20 soybean linkage groups (Song et al. 2004) for a background selection purpose and their leaflet shape trait was assessed by visual inspection (J.-K. Moon, unpublished data). Finally, the Sowon plants were crossed with a BC_2F_1 plant that was found to harbor the homozygous broad leaflet trait and *Rsv*4 gene based on the analysis of their progeny. Four families of F_2 plants derived from four BC_3F_1 seeds were segregated for both narrow/broad leaflet shapes and *Rsv*4/*rsv*4. The four families, which contained 60 (PM340), 64 (PM341), 72 (PM342) and 113 (PM343) F_2 plants, respectively, were combined into the SV population, which consisted of a total of 309 F_2 individuals. The F_2 seeds were spaced at 15-cm intervals in consecutive hill rows separated by 1 m in the

Ochang field of the Korea Research Institute of Bioscience and Biotechnology in 2007.

Measurement of leaflet shape

The leaflet shape of each of the F_2 plants was determined on the basis of the leaflet length/width ratio of the F_2 plant itself, as well as 15–20 F_3 plants from each of 309 $F_{2:3}$ lines, as recommended by previous studies (Sawada 1992; Chen and Nelson 2004). The selected representative leaflet of each F_2 plant was the center leaflet of a fully expanded leaf at the third or fourth node on the main stem from the top of the reproductive phase 1 plant (Fehr and Caviness 1977). Fifteen seeds from each $F_{2:3}$ line were planted in 18-cm plastic pots containing a 1:1 mixture of top soil and commercial potting soil in a greenhouse in the winter of 2007/08. The length/width ratio is the ratio between the maximum length to width of the leaflet. As the selected leaflet might not be a representative of the plant, additional visual inspections, including occasional measurements, of the shapes of many leaflets in each plant were conducted for verification purposes at each week over the next 3 weeks. The leaflet shapes were recorded as either narrow or broad.

Measurement of number of seeds per pod

When the plant growth reached reproductive phase 8, in which 95% of the pods had reached their mature (brown or tan) color, the numbers of 1-, 2-, 3-, and 4-seeded pods were recorded for each of the plants. The number of potential seeds for each plant was obtained by summing the numbers of 1-, 2-, 3-, and 4-seeded pods multiplied by one, two, three, and four, respectively. Pods having seeds that did not develop perfectly and even the pods that dropped were all included in the total number calculated. The seeds-per-pod value for each plant was calculated for each plant by dividing the total number of potential seeds by the total number of pods.

In order to precisely determine the exact genotype (in particular, heterozygous genotype) of each F_2 plant, the seeds-per-pod phenotypes were collected from 15 F_3 plants from each of 309 $F_{2:3}$ lines grown at the Ochang field in 2008. Fifteen seeds from each $F_{2:3}$ line were planted in the same fashion as were the F_2 seeds. The ends of the rows were bordered by 10 seeds of each of two parents. Although several poorly grown $F_{2:3}$ lines posed some difficulty, visual inspections almost always allowed us to determine the seeds-per-pod phenotypes of the $F_{2:3}$ lines. The phenotyping of those poor-grown $F_{2:3}$ lines and the lines that contained a recombination within a distance of approximately 1 centiMorgan (cM) from the *ln* locus was repeated in 2009.

DNA extraction and marker analysis

Young trifoliolate leaf tissues were collected from the 309 field-grown F_2 individuals of the SV population and soybean parents. Soybean genomic DNA was isolated as described previously by Saghai-Maroof et al. (1984). For quick preparation from the $F_{2:3}$ line plants, soybean genomic DNA was isolated using a FastDNA® Kit in accordance with the manufacturer's protocols (MP Biomedicals, Solon, OH, USA).

DNA from the SV population was genotyped using publicly available markers (Cregan et al. 1999; Song et al. 2004; Yang et al. 2008), as well as primers designed for this study, which were derived from the microsatellite sequences observed in the Williams 82 whole genome shotgun (WGS) sequence (Schmutz et al. 2010; also released at http://www.phytozome.net). For the markers designed in this study, SSR regions were identified via visual inspection of microsatellite repeats on the targeted sequence region using the 'Search' option in a text format, and primers were designed using the web-based Primer3 platform (Rozen and Skaletskyv 2000).

Amplification using microsatellite and allele-specific PCR primer sets was conducted as described by Jeong and Saghai Maroof (2004) and Cregan et al. (1999). In brief, the polymerase chain reaction (PCR) mixture contained 20 ng of total genomic DNA, 1× PCR buffer (10 mM Tris–HCl, 50 mM KCl, pH 8.3), 2.5 mM $MgCl_2$, 100 nM of each forward/reverse primer, 0.16 mM of each dNTP and 0.25 unit of *Taq* polymerase for a total volume of 20 µl. Standard PCR was conducted as follows: a denaturation step at 94°C for 5 min, 34 cycles at 94°C for 30 s, 43–58°C for 30 s and 72°C for 30 s and an extension step at 72°C for 5 min and followed by a 4°C soak. The PCR products were separated on either 6.25% denatured polyacrylamide gel electrophoresis (PAGE) with silver staining or on 2–3% agarose gel electrophoresis with ethidium bromide staining.

Heredity and linkage analysis

Chi-square analyses were performed to test for the goodness of fit to the expected ratios of the allele frequencies of leaflet shape trait, NSPP trait, and molecular marker data. Normality was evaluated using the QQ-plot computation in Excel (Bremer and Doerge 2010). The associations between markers and NSPP and leaflet length/width ratio traits were evaluated using the single-factor analysis of variance of the SAS® software version 9.2 (SAS Institute, Cary, NC, USA). Genetic map distances were calculated using the MAPMAKER 3.0b computer program (Lander et al. 1987), at a log likelihood of 5.0, with a maximum haldane distance of 50 cM.

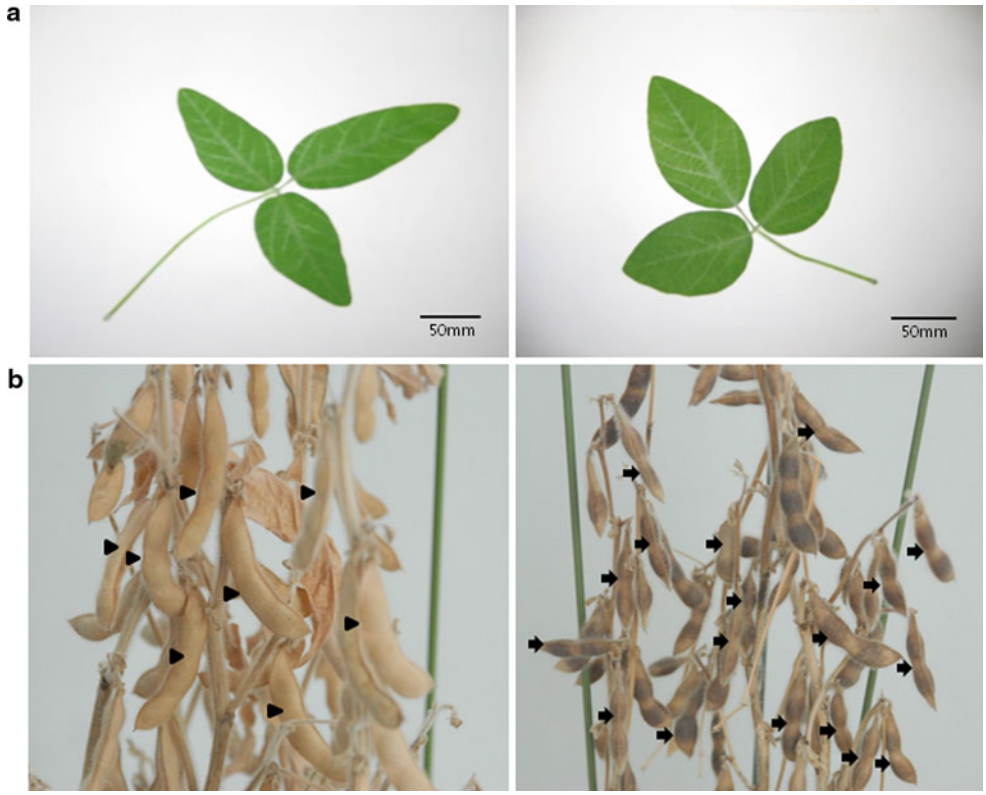

Fig. 1 The differential leaflet and pod shapes of the individuals of the BC_3F_2 population derived from the cross 'Sowon' × 'V94-5152'. **a** Representative trifoliolate leaves that show *narrow* leaflet shape in the Sowon type (*left*) and *broad* leaflet shape in the V94-5152 type (*right*). **b** Representative *pod* shapes that exhibit many more 3-seeded pods (marked with *arrow heads*) in the Sowon type and many more 2-seeded pods (marked with *arrows*) in the V94-5152 type

Results

Genetic segregation of leaflet shape

The results of measurements of the length/width ratios of the representative leaflets allowed the precise leaflet-shape phenotyping of most F_2 plants, with a few intermediate representative leaflets. However, subsequent visual observations of the shapes of many leaflets in each plant over 3 weeks allowed us to unambiguously distinguish between the broad and narrow leaflet lines in the SV population, thereby indicating that those intermediate leaflets represented atypical ones on those plants. Thus, the F_2 plants from the SV population could be divided into broad or narrow leaflet lines on the basis of the leaflet length/width ratios of the selected representative leaflets (Fig. 1a), in addition to visual observations conducted over the 3 weeks of the study. The mean value of the narrow leaflet type for the length/width ratio among F_2 plants from the SV population was 2.62 ± 0.2 (Table 1). The mean length/width ratio of the broad leaflet type was 1.84 ± 0.19. Thus, the difference of the mean values between the two leaflet types is comparable to the difference of the leaflet length/width ratios between Sowon and V94-5152 (Table 1). The leaflet shape of the Sowon (*ln*) type was categorized as the narrow (lanceolate) class and the leaflet shape of the V94-5152 (*Ln*) type was categorized as the broad (ovate) class. The segregation of the leaflet shapes of F_2 plants fitted a ratio of 3 broad:1 narrow (Table 2). The segregation of the leaflet shapes of F_2 plants obtained from 15 F_3 plants from each of 309 $F_{2:3}$ lines displayed a 1:2:1 ratio (broad leaflet: heterozygous: narrow leaflet) (Table 2).

Genetic segregation of the number of seeds per pod

Initially, we attempted to phenotype the seeds-per-pod trait of the F_2 plants in the PM340 and PM343 populations, on the basis of the presence/absence of the 4-seeded pod, as suggested previously by Zhu and Sun (2006), assuming that the 4-seeded pod would exist predominantly in the narrow leaflet-type plants. The segregation ratio of absence to presence of the 4-seeded pod appeared to fit a ratio of 1:3 (Table 2). These results seemed to indicate that, in contrast to the widely accepted hypothesis that both the seeds-per-pod and leaflet-shape traits are recessive, the presence of the 4-seeded pod was governed by a single dominant locus. When the obtained phenotype data were compared with the leaflet shape data in addition to the molecular marker data

Table 1 Means of Sowon and V94-5152 plants for leaflet- and seed-related traits combined over 2 years, 2007 and 2009

Soybean parent	Leaflet shape	Leaflet length/width ratio	Number of seeds per pod	Number of seeds per plant[a]	Percentage of 1-, 2-, 3, and 4-seeded pods				Weight of 100 seed (g)
					1-seeded	2-seeded	3-seeded	4-seeded	
Sowon	Narrow	2.93	2.73	707	3	19	71	7	13.87
V94-5152	Broad	1.93	1.98	687	13	78	9	0	16.36

[a] Except the number of seeds per plant, all other trait means were significantly different at the 0.05 significance level between Sowon and V94-5152 in paired *t* test analysis

Table 2 Segregation and chi-square tests for leaflet length/width ratios, numbers of seeds per pod, and presence/absence of 4-seeded pods of BC_3F_2 and $BC_3F_{2:3}$ progeny from Sowon × V94-5152

Trait[a]	Generation	Observed number of plants			Expected ratio	Goodness-of-fit	
		V94-5152 type	Heterozygote	Sowon type		χ^2	*P*
Leaflet length/width ratio	F_2	241	–	68	3:1	1.48	0.78
	$F_{2:3}$	78	163	68	1:2:1	1.58	0.55
Number of seeds per pod	F_2	132		38	3:1	0.635	0.58
	$F_{2:3}$	78	163	68	1:2:1	1.58	0.55
Presence/absence of 4-seeded pods	F_2	32	–	139	1:3	3.60	0.06

[a] For leaflet length/width ratio, V94-5152 type is broad and Sowon type narrow; for number of seeds per pod, V94-5152 type is low and Sowon type high; for presence/absence of 4-seeded pods, V94-5152 type is absence and Sowon type presence

(described below), approximately two-thirds of the F_2 individuals with the 4-seeded pods comprised the heterozygous leaflet-shape plants with a majority of 2-seeded pods. The 4-seeded pods were also observed in 22 of the homozygous broad leaflet-shape plants. Regardless of this observation, because only one or two of the 4-seeded pods were detected in many heterozygous individuals, correct phenotyping of the seeds-per-pod trait solely on the basis of the presence/absence of the 4-seeded pod was clearly problematic. Therefore, the seeds-per-pod trait was phenotyped based on the number of seeds per plant divided by the number of pods per plant rather than the presence/absence of 4-seeded pod, as suggested previously by Domingo (1945) and Dinkins et al. (2002).

The seed traits of the Sowon cultivar was compared with those of V94-5152 (Table 1). The Sowon plants contained 3% one-, 19% two-, 71% three-, and 7% 4-seeded pods. V94-5152 contained 13% one-, 78% two-, 9% three-, and 0% 4-seeded pods. Thus, the 3-seeded pod was identified as the principal seeds-per-pod characteristic of the Sowon cultivar and the 2-seeded pod was the major characteristic of the V94-5152 variant (Fig. 1b). However, because V94-5152 contained a far greater number of 2-seeded pods than the Sowon cultivar, the total numbers of seeds per plant of the Sowon cultivar was not significantly different from that of the V94-5152 cultivar. The weight per 100 seeds of V94-5152 was greater than that of Sowon. Thus, yield analyses including the total seed weight per unit area or the total number of seeds per plant, which were reported by Mandl and Buss (1981) using the broad and narrow leaflet soybean isolines, could not be applied to the SV population.

The number of seeds per pod and the number of pods of 170 F_2 individuals in the PM340 and PM343 subpopulations of the SV population were thoroughly counted (Fig. 2). A significant and positive correlation was found to exist between the NSPP values and the length/width ratios of the representative leaflets of the 170 F_2 individuals (Pearson's correlation coefficient = 0.64, $P < 0.0001$). The frequency distributions of the NSPP values for each plant in those subpopulations are shown in Fig. 2. The distribution tended toward bimodality with high values constituting the smaller portion of the curve, as would be expected in cases of single dominant gene inheritance. When these data were divided by the genotypes of the leaflet-shape trait already established for the F_2 individuals, the frequency of each genotype was distributed normally. The two unimodal distributions appeared to be normal distributions on the normality tests. The modal distribution with high NSPP values contained homozygous narrow leaflet individuals, and the modal distribution with low values harbored homozygous and heterozygous broad leaflet individuals, with a single exception. The NSPP value of the single exceptional individual was rather low, owing to the unusually high number of 1-seeded pods. Although the average number of 1-seeded pods for a plant was 13.5 in the total population, the plant contained 76 one-seeded, 120 two-seeded, 143 three-seeded, and 15 four-seeded pods. The NSPP value of the V94-5152 was located at the lower end of the modal

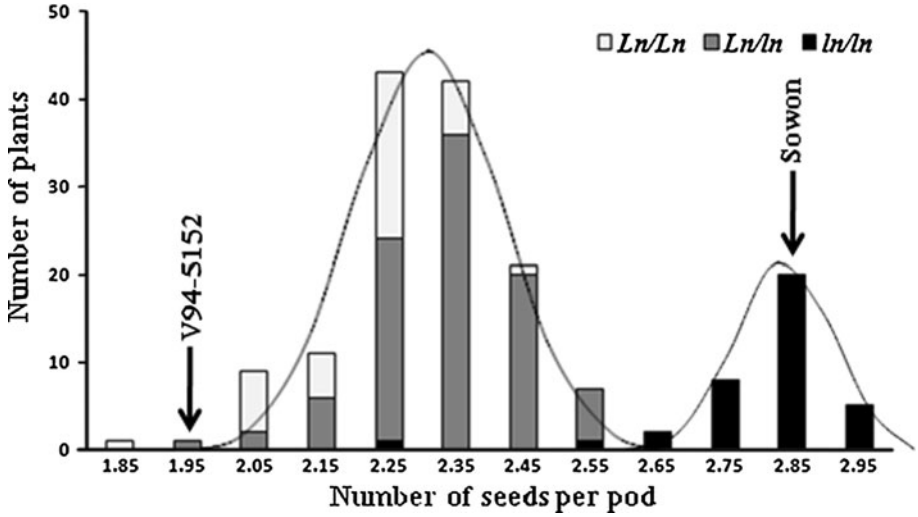

Fig. 2 Bimodal distributions of the number of seeds per pod (NSPP) for a total of 170 F_2 individuals in the 'Sowon' × 'V94-5152' population. The NSPP values were normally distributed when they were divided by *Ln/Ln* and *Ln/ln* genotype and *ln/ln* genotype groups

distribution with low NSPP values, thereby suggesting the presence of additional modifying genes likely from the Sowon cultivar. The means and standard deviations of the larger and smaller normal distributions were 2.31 ± 0.12 and 2.78 ± 0.21, respectively. The number of segregates in the larger and smaller normal distributions showed a good fit to a 3:1 ratio (Table 2).

The NSPP phenotypes of the rest of the F_2 individuals in the SV population, as well as those from each of the 309 $F_{2:3}$ lines, were determined via visual inspection. The segregation for the NSPP genotypes of F_2 plants obtained from the 15 F_3 plants derived from each of 309 $F_{2:3}$ lines displayed a 1:2:1 ratio (Table 2). Comparisons between the leaflet-shape and NSPP genotypes of each of the 309 F_2 plants in the SV population demonstrated that, in every case, the broad leaflet-shape individuals corresponded with the low NSPP individuals, the leaflet-shape heterozygotes with the NSPP heterozygotes, and the narrow leaflet-shape individuals with the high NSPP individuals. The results demonstrated that the two traits cosegregated in the SV population.

Linkage mapping of *ln*

The genotyping results of 309 F_2 individuals of the leaflet shape were cosegregating with those of the NSPP trait determined from the $F_{2:3}$ lines. Because our mapping population was an advanced backcrossed BC_3F_2 population, 6.25% of the polymorphic markers between Sowon and V94-5152 would be hypothesized to be polymorphic in the BC_3F_2 population. In other words, a portion of the polymorphic markers that were originally polymorphic between Sowon and V94-5152 ought to be linked to the leaflet shape and NSPP traits segregating in the BC_3F_2 population. Therefore, we attempted to screen the public soybean microsatellite markers reported in the studies of Cregan et al. (1999), Song et al. (2004), and Yang et al. (2008), as part of a broader effort to select polymorphic markers between the Sowon, BC_2F_1 obtained by pooling genomic DNAs from the BC_2F_2 plants (progeny of the BC_2F_1 donor parent harboring the homozygous *Ln* gene on the basis of the phenotyping of its F_2 progeny), and V94-5152 plants. At the time at which the nearly 50 microsatellite markers were screened, Sat_105 and Satt270 on the soybean molecular linkage group I (chromosome 20) were found to be polymorphic between the BC_2F_1 and Sowon plants. Interestingly, Sat_105 and Satt270 have been previously identified as markers linked to the *ln* gene (Song et al. 2004). Then, Sat_105 and Satt270 were genotyped in the 309 individuals of the SV population. As a result, Sat_105 and Satt270 were mapped 3.1 and 3.6 cM from the *ln* locus, respectively, on one side of *ln* in the SV population (Fig. 3). Single-factor analysis of variance also identified significant associations ($P < 0.0001$) between these markers and the quantitative NSPP values and leaflet length/width ratios of the F_2 individuals in the SV population (Table 3). To further corroborate the chromosomal region on chromosome 20 and also to map markers on the other side from *ln*, additional public markers, which were expected to delimit the *ln* locus detected by Sat_105 and Satt270 on the basis of the maps constructed by Song et al. (2004) and Yang et al. (2008), were mapped in the SV population. We mapped a total of three public microsatellite markers (SM315, Satt270, and Sat_105) to one side of the vicinity of the *ln* locus and mapped the four public microsatellite markers (Sat_268, Satt049, Satt671, and Sat_418) to the other side of *ln* (Fig. 3). The results showed that the narrow leaflet-determining gene in Sowon is an allele of the previously established *ln* gene.

Construction of the high-resolution map

The sequences of the public molecular markers delimiting *ln* were used in BLAST searches against the soybean whole-genome sequence. The results demonstrated that,

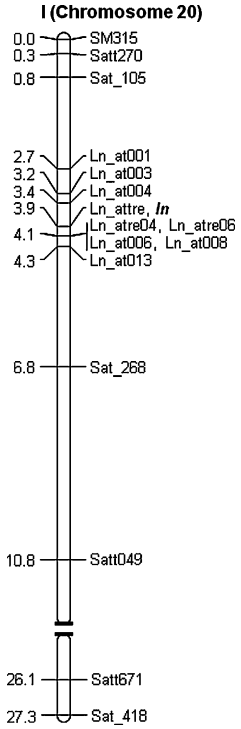

Fig. 3 High-resolution genetic map of the chromosomal region around the *ln* on molecular linkage group I (chromosome 20), encoding for a leaflet shape regulator. Markers were mapped in a 309-individual BC_3F_2 population of the 'Sowon' (*ln*) × 'V94-5152' (*Ln*). Map distances of markers from SM315 are given in centiMorgans (cM). For this particular genetic map, a genetic distance of 0.2 cM represents a single recombination event between two loci

based on the positive hits containing the most significant alignment, Sat_105 and Sat_268, the closest markers on both sides of *ln*, delimit 0.94 Mb between the nucleotide positions 34.23 Mb (Sat_105) to 35.17 Mb (Sat_268) on chromosome 20 of the soybean plant.

In order to develop markers linked more closely to the *ln* gene, primers were designed to target microsatellite regions in this chromosome sequence (Table 4). As a result, we mapped nine new microsatellite markers in the vicinity of *ln* (Fig. 3). Ln_atre04, Ln_atre06, Ln_at006, and Ln_at008 were cosegregating in our mapping population; however, Ln_atre04 is the marker located closest to *ln* on the soybean chromosome sequence. In our mapping population, a genetic distance of 0.2 cM represented a single recombination event between two loci. Among the markers, Ln_at004, the closest markers to *ln* on the Sat_105 side, is located 0.5 cM from *ln*; Ln_atre04, the closest marker on the Sat_268 side, is located 0.2 cM from *ln*; and Ln_attre cosegregates with *ln*. Single-factor analysis of variance also identified significant associations ($P < 0.0001$) between these markers and the quantitative NSPP values and leaflet length/width ratios of the F_2 individuals in the SV population (Table 3). The markers accounted for approximately 30% more phenotypic variation for the leaflet length/width ratio trait than for the NSPP trait. Ln_attre, which cosegregates with *ln*, showed the highest R^2 values for both the NSPP (55.4%) and leaflet length/width ratio (83.7%) traits (Table 3). Mapping the two markers Ln_at004 and Ln_atre04 to the soybean whole genome shotgun sequence (Schmutz et al. 2010), we determined that this 0.7-cM region corresponded to a sequence length of approximately 66 kb, giving rise to a recombination rate of 93 kb/cM. This recombination rate is approximately 4.5-fold higher than

Table 3 Single-factor analysis of variance between the leaflet length–width ratio and number of seeds per pod and markers mapped between Satt270 and Sat_268 on soybean chromosome 20 in a BC_3F_2 population from the cross of Sowon and V94-5152

Trait	Marker	*P*	R^2 (%)	Allelic means[a]		
				SS	SV	VV
Leaflet length/width ratio	Satt270	<0.0001	68.5	2.56 (67)	1.92 (166)	1.66 (76)
	Sat_105	<0.0001	70.1	2.56 (68)	1.96 (163)	1.66 (78)
	Ln_at001	<0.0001	77.3	2.59 (69)	1.95 (162)	1.65 (78)
	Ln_at003	<0.0001	79.1	2.59 (70)	1.94 (161)	1.65 (78)
	Ln_at004	<0.0001	80.3	2.60 (69)	1.94 (162)	1.65 (78)
	Ln_attre	<0.0001	83.6	2.62 (68)	1.94 (163)	1.65 (78)
	Ln_atre04, Ln_atre06, Ln_at006, Ln_at008	<0.0001	83.6	2.62 (68)	1.94 (164)	1.64 (77)
	Ln_at013	<0.0001	82.6	2.61 (69)	1.94 (163)	1.64 (77)
	Sat_268	<0.0001	74.1	2.58 (70)	1.94 (163)	1.65 (73)
Number of seeds per pod	Satt270	<0.0001	43.6	2.71 (37)	2.37 (95)	2.22 (39)
	Sat_105	0.0089	45.5	2.71 (38)	2.36 (94)	2.22 (39)
	Ln_at001	<0.0001	48.3	2.81 (37)	2.34 (96)	2.22 (38)
	Ln_at003	<0.0001	50.9	2.73 (38)	2.35 (95)	2.22 (38)
	Ln_at004	<0.0001	50.9	2.73 (38)	2.35 (95)	2.22 (38)
	Ln_attre	<0.0001	55.4	2.75 (38)	2.35 (95)	2.22 (38)
	Ln_atre04, Ln_atre06, Ln_at006, Ln_at008	<0.0001	55.4	2.75 (38)	2.35 (95)	2.22 (38)
	Ln_at013	<0.0001	55.4	2.75 (38)	2.35 (95)	2.22 (38)
	Sat_268	<0.0001	53.0	2.74 (38)	2.35 (90)	2.23 (39)

Numbers of F_2 individuals are presented in parentheses

[a] *SS* homozygous Sowon, *VV* homozygous V94-5152, *SV* heterozygous

Table 4 List of PCR primer sequences, core repeat motifs, and nucleotide positions on the soybean whole genome shotgun sequence of microsatellite markers linked to the *ln* locus developed in this study

Marker name	Primer specificity	Primer sequence (5′ → 3′)	Core repeat motif	Nucleotide position on chromosome 20 (kb)
Ln_at001	Forward	CAAGATACACTTAGCCTTTT	AT	34586
	Reverse	TTATATTCAATCCTTATCGTG		
Ln_at003	Forward	TTCTATCGATTTGTCATTTT	AT	34616
	Reverse	TTATGTCTTAGTATAAGAATTTGA		
Ln_at004	Forward	AAAAGAATATATATCTGTGATGA	AT	34638
	Reverse	TATTTATGCTGTTTCATAAT		
Ln_attre	Forward	GTTTCTTTTTGCCATTTAC	ATT	34687
	Reverse	GAATATGCAGAAGCAATTAAC		
Ln_atre04	Forward	GATTTGATCTCTTTTCATTCA	AT	34704
	Reverse	GAAATCCAACTACGTACACTG		
Ln_atre06	Forward	GAATCATGAATATGGGAATA	AT	34709
	Reverse	GCAAATCCACCATACATACAT		
Ln_at006	Forward	AAATGATAATCGAATTGCT	AT	34726
	Reverse	TTCAAAAACATTTTCCTTTA		
Ln_at008	Forward	AGTTGTGAAGTGTGTTCAAT	AT	34750
	Reverse	TAAATGAAAATTTCGGATAA		
Ln_at013	Forward	AAAAATATCAAAGGGAGATT	AT	34842
	Reverse	TTATGCTAGTTGCTTCTTTT		

the genome average of approximately 440 kb/cM. We also observed two ambiguous sequence components presented by N in this 66-kb region. Thus, the exact size of the DNA fragment between Ln_at004 and Ln_atre04 could not be predicted.

According to the soybean gene annotation database accessible at Phytozome v5.0 (http://www.phytozome.net, as of March 2010), only three predicted genes (Glyma20g 24980.1, Glyma20g24990.1, and Glyma20g25000.1) exist within this 66-kb region. Glyma20g24980.1 may encode for a hypothetical protein, Glyma20g24990.1 for a ribosomal protein L15, and Glyma20g25000.1 for another hypothetical protein. Of particular interest, the predicted peptide sequence of Glyma20g25000.1 is 47.8% homologous to the zinc finger (C2H2 type) family protein encoded for by the *JAGGED* gene, which is necessary for the proper shaping of lateral organs and is sufficient to induce the proliferation of lateral organ tissue (Dinneny et al. 2004; Ohno et al. 2004).

Discussion

The first step in establishing a marker-assisted selection scheme for leaflet shape-regulating genes would be to construct a high-resolution genetic linkage map in the vicinity of the *ln* gene. Sat_105 and Sat_268 have been previously identified as markers linked to the *ln* gene (Song et al. 2004). Our results demonstrated that the narrow leaflet-determining gene in the Sowon cultivar is located between Sat_105 and Sat_268, thereby indicating that the gene in Sowon is an allele of the previously highly studied *ln* gene. We then constructed a high-resolution map by locating additional markers developed from the soybean genome sequence (Schmutz et al. 2010). The high-resolution map delimited the *Ln* chromosomal region to a sequence length of approximately 66 kb. The results showed that one of the predicted genes, which is 47.8% homologous to the *JAGGED* gene that regulates lateral organ development (Dinneny et al. 2004; Ohno et al. 2004), is located within the same chromosomal region as the *Ln* locus. Although further experimental tests including a complementation test will be required in the near future, the gene is likely a good candidate gene regulating the broad/narrow leaflet shape, as well as the NSPP trait.

Phenotyping methods of the number-of-seeds-pod trait have been inconsistent in previous studies. Takahashi (1934) divided the phenotypes of the NSPP trait into two classes, according to whether less than 10% or more than 10% of the pods from a given plant were 4-seeded. More recently, Zhu and Sun (2006) divided the F_2 plants into two classes based on the presence/absence of the 4-seeded pod. By contrast, Domingo (1945) and, more recently, Dinkins et al. (2002) divided the phenotypes of the NSPP trait into two classes, using the values obtained by dividing the total number of potential seeds by the total number of pods. In

our initial attempt to assess the phenotype of the NSPP trait, we followed the former, simpler method. The comparison of the obtained 4-seeded-pod phenotype data with the leaflet shape and molecular marker data demonstrated that, in addition to plants with a homozygous narrow leaflet shape (*ln*/*ln*), the majority of the plants heterozygous in the leaflet shape (*Ln*/*ln*) and some of the homozygous broad leaflet-shape (*Ln*/*Ln*) plants with the low NSPP values contain more than one 4-seeded pod. The results indicated that the presence of the 4-seeded pod predominated in our mapping population. By way of contrast, when we compared the leaflet shape genotypes with the genotypes of the NSPP trait obtained via the latter method, the high NSPP phenotype appeared to be recessive and both the leaflet shape and the NSPP genotypes cosegregated in our mapping population. Our findings suggested that, consistent with the conclusions by Weiss (1970), the seeds-per-pod trait, or at least the presence/absence of the 4-seeded pod, is governed not only by a single *ln* or a gene tightly linked with *ln*, but also by additional modifying gene(s) that are probably present in the Sowon cultivar. This modifying gene hypothesis is supported to some degree by the location of the NSPP value of the V94-5152 at the lower end of the modal distribution with low NSPP values (Fig. 2), the poor fit of the presence/absence of the 4-seeded pod to a 3:1 ratio (Table 2), and the lower explained phenotypic variation for the NSPP trait than for the leaflet length/width ratio trait observed in the single-marker analysis (Table 3).

A long-debated hypothesis in soybean research involves whether the leaflet trait (narrow or broad) and the NSPP trait are determined by independent genes, or represent pleiotropic effects of a single gene. This study demonstrated that these traits cosegregate in the Sowon × V94-5152 BC_3F_2 population. In flowering plants, new lateral organs including leaves and flowers develop continuously via a reiterative organogenesis process that occurs at the shoot apical meristem. Leaf development mutations frequently induce pleiotropic phenotypes and this phenomenon is particularly salient in the process of flower development (Johnson and Bernard 1962; Weiss 1970; Hofer et al. 1997; Dinneny et al. 2004; Ohno et al. 2004; Tattersall et al. 2005). For examples, *LEAFY/UNIFLOLIATA* regulates both leaf and flower morphogenesis in the pea (Hofer et al. 1997). The pea mutant *crispa*, which is defective in the *PHANTASTICA* gene and evidences a reduced leaflet width-to-length ratio similar to that of the soybean *ln* mutant, exhibited pleiotropic phenotypes in several aspects of flower development (Tattersall et al. 2005). Conversely, combinations of floral homeotic mutations result in the conversion of floral organs to leaf-like structures (Bowman et al. 1993). These results indicated that common regulatory processes are operant during the production of leaves and flowers. Interestingly, seed production per pod, one of the relevant characteristics of mature flowers, was lower in *cri-2* mutants than in the the wild-type cultivar, due to lower numbers of ovules and less frequent fertilization in *cri-2* mutant carpels than in the wild-type cultivar (Tattersall et al. 2005). In this study, we demonstrated that, unlike the *crispa* mutant, the *ln* mutant most likely induces a pleiotropic phenotype, increasing the NSPP, with a lack of complementation test.

Obtaining greater insights into the inheritance of the narrow leaflet trait may prove beneficial in breeding efforts for specialty small-seeded soybeans, including soybean sprouts (Lee et al. 2001). We have constructed a high-resolution map of a chromosomal region controlling the narrow leaflet trait for the eventual cloning of the *ln* gene, thus suggesting that yield, like other traits, may be improved by individually manipulating the component traits using molecular genetic tools. This research provides a leading example of the exploitation of the soybean whole genome sequence to elucidate an economically important soybean gene at the molecular level. The molecular markers that resulted from this research are expected to benefit soybean breeders in a fairly direct way. The future cloning and functional elucidation of this yield-regulating locus is expected to provide fundamental knowledge useful in efforts to improve the yields of soybean and other crops.

Acknowledgments This work was supported mainly by a Grant (Code PJ0070352010) from the BioGreen 21 Program, Rural Development Administration, and partly by the Korea Research Institute of Bioscience and Biotechnology Research Initiative Program and by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST). We thank two anonymous reviewers for critical reading and suggestions on how to improve the manuscript.

References

Bernard RL, Weiss MG (1973) Qualitative genetics. In: Caldwell BE (ed) Soybeans: improvement, production and uses. American Society of Agronomy, Madison, pp 117–154

Bowman JL, Alvarez J, Weigel D, Meyerowitz EM, Smyth DR (1993) Control of flower development in *Arabidopsis thaliana* by *APETALA1* and interacting genes. Development 119:721–743

Bremer M, Doerge RW (2010) Statistics at the bench: a step-by-step handbook for biologists. Cold Spring Harbor Laboratory Press, Cold Spring Harbor

Buss GR, Ma G, Chen P, Tolin SA (1997) Registration of V94-5152 soybean germplasm resistant to soybean mosaic potyvirus. Crop Sci 37:1987–1988

Chen Y, Nelson RL (2004) Evaluation and classification of leaflet shape and size in wild soybean. Crop Sci 44:671–677

Cregan PB, Jarvik J, Bush AL, Shoemaker RC, Kahler LarkKG, AL KayaN, VanToai TT, Lohnes DG, Chung J, Specht JE (1999) An integrated genetic linkage map of the soybean genome. Crop Sci 39:1464–1490

Dinkins RD, Keim KR, Farno L, Edwards LH (2002) Expression of the narrow leaflet gene for yield and agronomic traits in soybean. J Hered 93:346–351

Dinneny JR, Yadegari R, Fischer RL, Yanofsky MF, Weigel D (2004) The role of *JAGGED* in shaping lateral organs. Development 131:1101–1110

Domingo WE (1945) Inheritance of number of seeds per pod and leaflet shape in the soybean. J Agric Res 70:251–268

Fehr WR, Caviness CE (1977) Stages of soybean development, Iowa State University, Special Report 80. Iowa State University, Ames

Hofer J, Turner L, Hellens R, Ambrose M, Matthews P, Michael A, Ellis N (1997) *UNIFOLIATA* regulates leaf and flower morphogenesis in pea. Curr Biol 7:581–587

Hwang TY, Moon J-K, Yu S, Yang K, Mohankumar S, Yu YH, Lee YH, Kim HS, Kim HM, Saghai Maroof MA, Jeong SC (2006) Application of comparative genomics in developing molecular markers tightly linked to the virus resistance gene *Rsv*4 in soybean. Genome 49:380–388

Jeong SC, Saghai Maroof MA (2004) Detection and genotyping of SNPs tightly linked to two disease resistance loci, *Rsv*1 and *Rsv*3, of soybean. Plant Breed 123:305–310

Johnson HW, Bernard RL (1962) Soybean genetics and breeding. Adv Agron 14:149–221

Kwon SH, Lee YI, Kim JR (1981) Evaluation of important sprouting characteristics of edible soybean sprout cultivars. Korean J Breed 13:202–206

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, 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, Park KY, Lee HS, Park EH, Boerma HR (2001) Genetic mapping of QTLs conditioning soybean sprout yield and quality. Theor Appl Genet 103:702–709

Mandl FA, Buss GR (1981) Comparison of narrow and broad leaflet isolines of soybean. Crop Sci 21:25–27

Ohno CK, Reddy GV, Heisler MGB, Meyerowitz EM (2004) The Arabidipsis *JAGGED* gene encodes a zinc finger protein that promotes leaf tissue development. Development 131:1111–1122

Park KY, Yun HT, Moon J-K, Ku JH, Hwang JJ, Lee SH, Seung YK, Ryu YH, Chung WK, Lee YH, Kim SD, Chung MN (2000) A new soybean cultivar for sprout with good storability and disease resistance, 'Sowonkong'. Kor J Breed 32:298–299

Pedersen P, Lauer JG (2004) Response of soybean yield components to management system and plant date. Agron J 96:1372–1381

Rozen S, Skaletskyv HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics methods and protocols: methods in molecular biology. Humana Press, Totowa, pp 365–386

Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci USA 81:8014–8081

Sawada S (1988) Inheritance of leaflet shape in soybeans. Soybean Genet Newsl 15:61–65

Sawada S (1992) Time of determination and variations within and between plants in leaf shape of soybean. Jpn J Crop Sci 61:96–100

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

Song QJ, Marda 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

Takahashi N (1934) Linkage relation between the genes for the form of leaves and the number of seeds per pod of soybeans. Jpn J Genet 9:208–225

Tattersall AD, Turner L, Knox MR, Ambrose MJ, Ellis THN, Hofer JMI (2005) The mutant *crispa* reveals multiple roles for *PHANTASTICA* in pea compound leaf development. Plant Cell 17:1046–1060

Weiss MG (1970) Genetic linkage in soybeans. Linkage group IV. Crop Sci 10:368–370

Yang K, Moon J-K, Jeong N, Back K, Kim HM, Jeong SC (2008) Genome structure in soybean revealed by a genomewide genetic map constructed from a single population. Genomics 92:52–59

You MG, Liu YB, Zhao TJ, Gai JY (1995) Effects of leaf shape on seed yield and its components in soybeans. Soybean Genet Newsl 22:66–70

Zhu BG, Sun YR (2006) Inheritance of the four-seeded-pod trait in a soybean mutant and marker-assisted selection for this trait. Plant Breed 125:405–407