Quantitative Trait Loci for β-Conglycinin (7S) and Glycinin (11S) Fractions of Soybean Storage Protein

D.R. Panthee[a], P. Kwanyuen[b], C.E. Sams[a], D.R. West[a], A.M. Saxton[c], and V.R. Pantalone[a,*]

[a]Department of Plant Sciences, University of Tennessee, Knoxville, Tennessee 37996, [b]USDA, ARS, Raleigh, North Carolina 27607, and [c]Department of Animal Science, University of Tennessee, Knoxville, Tennessee 37996

ABSTRACT: Glycinin (11S) and β-conglycinin (7S) are important seed storage proteins in soybean [*Glycine max* (L.) Merr.]. A major limitation of soybean seed storage proteins is their low levels of the sulfur-containing amino acids, methionine and cysteine, which are important nutritional components of protein meal. Glycinin contains significantly more S-containing amino acids than does β-conglycinin. Thus, detection of quantitative trait loci (QTL) that govern 11S may provide marker-assisted selection (MAS) opportunities to improve soybean total S-containing amino acids. The objective of this study was to detect and map QTL governing 7S and 11S fractions of soybean seed storage proteins. To achieve this objective, 101 F_6-derived recombinant inbred lines (RIL) developed from a cross of N87-984-16 × TN93-99 were used. Storage proteins were extracted from all RIL and separated in 10–20% linear gradient polyacrylamide gels. Dried gels were scanned for individual subunits of storage protein with a densitometer equipped with a He-Ne laser light source. Data were converted to concentration for each subunit component and analyzed using SAS software. A significant ($P < 0.05$) difference among genotypes was found for glycinin and β-conglycinin. A total of 94 polymorphic simple sequence repeat molecular genetic markers were used in screening all RIL. Three QTL for glycinin (Satt461, Satt292, and Satt156) were distributed on linkage group (LG) D2, I, and L, respectively, whereas two QTL for conglycinin (Satt461 and Satt249) were distributed on LG D2 and J. Phenotypic variation explained by individual QTL ranged from 9.5 to 22%. These QTL may provide useful MAS opportunities for improvement of nutritional quality in soybean.

Paper no. J10857 in *JAOCS 81*, 1005–1012 (November, 2004).

KEY WORDS: β-Conglycinin, *Glycine max*, glycinin, QTL mapping, soybean, storage protein.

Soybean is a major source of protein and amino acids for human and animal feeds in the world (1). However, soy protein is not perfect because of its low levels of the sulfur-containing essential amino acids, methionine and cysteine. Enhancing the concentration of nutritionally essential amino acids in soybean would improve soy meal quality and boost its inclusion in diets for humans and animals.

Soybean storage protein has two major fractions, β-conglycinin (7S) and glycinin (11S), accounting for more than 70% of the total proteins (2). β-Conglycinin is a trimer with subunits α, α′, and β, and a M.W. of about 180 kDa. Glycinin is a hexamer with a M.W. of 360 kDa, consisting of acidic and basic subunits. Glycinin constitutes about 35% of the total seed storage protein (3). Glycinin has three to four times more S-containing amino acids (particularly methionine) than does β-conglycinin (4). In addition, the β subunit of conglycinin is known to be void of methionine and cysteine. Hence, the glycinin fraction may be more desirable than the β-conglycinin fraction in developing improved amino acid balance in soy meal. Both 7S and 11S are components of total seed protein, and there is a typical inverse relationship between 7S and 11S concentrations (5). Thus, glycinin can be increased at the expense of β-conglycinin. Mutants with increased glycinin concentration have been generated using this concept by some plant breeders (e.g., 6,7).

Yaklich (8) studied the storage protein of high-protein lines of soybean to determine how the two major storage proteins contribute to the increased protein content. It was found that high seed protein lines contain more β-conglycinin and glycinin than normal protein lines. From these data, one can infer that quantitative differences exist in protein bands of major seed storage proteins, indicating that there should be differences in the genome that are responsible for the synthesis of different quantities of subunits of the storage protein. In fact, Nielsen *et al.* (9) identified five genes regulating the production of glycinin in soybean. Harada *et al.* (10) showed that three genes control the production of β-conglycinin. However, the position of those genes on the soybean genetic linkage map remains to be determined.

Hayashi *et al.* (11) used amplified fragment length polymorphism markers to locate the *cgdef* locus in soybean. This locus determines the β-conglycinin content in soybean, and a single recessive gene reduces the β-conglycinin content. Such reduction results in an increased level of glycinin. A mutant line (Keburi) with reduced β-conglycinin was also identified by Kitamura and Kaizuma (12).

*To whom correspondence should be addressed Department of Plant Sciences, University of Tennessee, 2431 Joe Johnson Dr., Knoxville, TN 37996. E-mail: vpantalo@utk.edu

Among the efforts to increase the S-containing amino acids in the soy protein fraction, genetic transformation was accomplished by introducing a gene from Brazil nut (*Bertholletia excelsa*), but the project was abandoned because of allergenicity issues resulting from the Brazil nut transgene (13). *Glycine soja*, the wild ancestor of soybean, also has been studied to increase glycinin; however, hybridizations between soybean and *G. soja* have not yet led to the development of an agriculturally acceptable cultivar (14,15).

In our experiment, the parental line N87-984-16 is one of two essentially equivalent F_8-derived sister lines that constitute the high protein commercial soybean cultivar 'Prolina' (16). Prolina has high seed protein concentration and also exhibits superior protein gelling properties, indicating that its storage proteins may contain more disulfide bonds and increased S-containing amino acids (17). The other parental line utilized in our experiment, TN93-99, is a high-yielding soybean line with typical protein concentration. TN93-99 is the registered germplasm GP-280 (18). The objective of the study was to detect and map quantitative trait loco (QTL) governing 7S and 11S fractions of the storage proteins in a soybean recombinant inbred line (RIL) population formed from the cross of N87-984-16 × TN93-99.

MATERIALS AND METHODS

Plant material. A total of 101 F_6-derived RIL were developed from a cross of N87-984-16 × TN93-99. The N87-984-16 line is one of two sister lines, whose blend constitutes the high-protein commercial cultivar 'Prolina' (16). The line TN93-99 is a high-yielding and well-adapted University of Tennessee breeding line and is currently registered as a germplasm (18). Crossing was made during summer of 1998. F_1 seeds were harvested in October 1998, and F_2 single plants were grown in Costa Rica during winter 1998. Generations were advanced until F_5 in Costa Rica and $F_{5:6}$ seeds were obtained in May 2000. About 300 $F_{5:6}$ single plants were grown at Knoxville as a RIL population, and 101 random plants of similar Tennessee-adapted maturity were selected as a source population for this study. A total of 101 F_6-derived RIL developed from a cross of N87-984-16 × TN93-99 were planted in a randomized complete block design with three replications at the Knoxville Experiment Station of the University of Tennessee (Knoxville, TN) in May 2001 and 2003. Each line was planted in plots of 6 m length with 75 cm between rows. In 2001, two-row plots were used, and four-row plots were used in 2003. Seeds were harvested with a two-row combine and stored below 10°C until analyzed.

Preparation of protein. Ten grams of soybean seeds were ground in a water-cooled (20°C) Knifetec 1095 Sample Mill (FOSS Tecator, Höganäs, Sweden) for 20 s. This setting produced soybean flour with a uniform particle size. Soluble protein was extracted for 1 h at room temperature with stirring from 1 g of full-fat soybean flour in a 1:15 (wt/vol) ratio with 0.2 M Tris-HCl buffer, pH 8.0, that contained 0.1 M β-mercaptoethanol. The mixture was centrifuged at 10,000 × *g* for 10 min at 4°C. After the fat layer was removed, an aliquot of 1 mL of crude protein extract or supernatant was taken from each sample. The total protein concentration of each sample was determined as described by Bradford (19). The modified protein assay kit was obtained from Bio-Rad (Richmond, CA). Storage proteins and their polypeptides were dissociated in the crude extract by adding an equal volume of 5% SDS and of 0.1 M β-mercaptoethanol solution to each sample and boiling in a water bath for 10 min for complete dissociation. Glycerol and bromophenol blue were added to each sample to achieve a final concentration of 10 and 0.025%, respectively.

Polyacrylamide protein gel electrophoresis. Proteins and their polypeptides were separated in a Bio-Rad Protean II vertical slab gel apparatus according to Chua (20) with the following modifications. Each sample, containing approximately 80 to 100 μg protein (approximately 10 μL of protein sample), was loaded onto the gel. Proteins and polypeptides were separated using a linear gradient of 10 to 20% polyacrylamide gel. The dimensions of the separating gel are 14 × 16 × 0.15 cm with 15 sample wells in the stacking gel. Blank sample wells were left between loaded samples to prevent protein cross-contamination during electrophoresis and to facilitate accurate quantification by scanning densitometry after electrophoresis. Since more than one gel was required for these protein samples, and since it was necessary to obtain the same polyacrylamide composition throughout the entire experiment, identical gradient gels were cast from the same polyacrylamide solution and from the same gradient-producing condition. Electrophoresis of each protein sample was carried out in duplicate at a constant current of 10 mA/gel at room temperature until the bromophenol blue or tracking dye reached the bottom of the gel.

Staining and destaining. Gels were fixed in 40% (vol/vol) methanol and 10% (vol/vol) acetic acid for at least half an hour prior to staining. They were stained in freshly prepared dye containing 0.25% Coomassie Brilliant Blue (wt/vol), 40% (vol/vol) methanol, and 10% (vol/vol) acetic acid on an orbit shaker with fixed speed. They were stained for at least 8 h. Gels were then destained in 40% (vol/vol) methanol and 10% (vol/vol) acetic acid on the same orbital shaker. The destaining solution was changed every 2 h for a minimum of four changes. The destaining was terminated when the background gels were almost visibly cleared of the dye. The residual dye on the gels did not interfere with the analysis of polypeptide bands since local average background subtraction was used in the analyses. Destained gels were soaked in deionized water for at least 5 min. Each gel was placed between two cellophane sheets and dried in a Bio-Rad GelAir dryer.

Scanning densitometry. Dried gels were scanned as described by Kwanyuen and Wilson (21) with a Molecular Dynamics Personal Densitometer SI (Sunnyvale, CA) equipped with a HeNe laser light source. ImageQuant (Molecular Dynamics, Sunnyvale, CA) software for volume integration was used in data analysis to determine total absorbance of entire protein bands (Fig. 1). The absorbance of protein on gels was

adjusted to ensure that the most intense protein bands were within the optimal and linear response range of the detector. Apparent absorbance of each protein was obtained by subtracting the background absorbance from the total absorbance of the protein bands within the same gel volume. The relative amount of each protein or polypeptide was expressed as a percentage of total protein in the same gel lane.

DNA extraction and polymerase chain reaction (PCR). DNA was extracted from the RIL and parental lines utilizing Qiagen Plant Easy DNA Extraction Kit (Qiagen, Hiden, Germany). PCR consisted of 7.4 μL of double-distilled H_2O, 1 μL of 10× PCR buffer, 1 μL of 2 mM dNTPs mixture (Pharmacia, Piscataway, NJ), 0.5 μL of 20 μM forward and reverse primer, 0.1 μL of 5 units/μL Klentaq (Ab Peptides Inc., St. Louis, MO), and 2 μL of 20 ng/μL template DNA. The PCR was performed in a 96-well MBS Hybaid thermocycler (Hybaid, Franklin, MA). PCR conditions were (i) 94°C for 5 min followed by 35 cycles at 94°C for denaturation for 25 s, (ii) 47°C for annealing for 30 s, (iii) 72°C for 25 s for extension, and one last cycle at 72°C for final extension for 5 min. Parents were screened for a total of 550 $(ATT)_n$ type of simple sequence repeat (SSR) genetic markers developed by Cregan *et al.* (22). The sequence information for the markers is publicly available from the USDA web site http://129.186.26.94/ssr.html, which was verified on April 12, 2004. A total of 94 SSR markers that were found to be polymorphic in the parents and among the RIL were screened and used in QTL analyses.

DNA gel electrophoresis. A 6% nondenaturing PAGE used to separate the PCR product consisted of 6% bis-acrylamide, 0.5% Tris borate EDTA buffer, 0.07% ammonium persulfate, and 0.035% tetramethylenediamine. Two microliters of loading buffer (6×) were added to PCR products, and a 10 μL sample was loaded on the gel. The running buffer was 0.5× TBE. The gel was run at a constant 300 V for 3 h. Circulated air was used to keep the glass plates cool during gel running. Ethidium bromide (50 μL, 10 mg/mL) was added to the running buffer to visualize the bands under exposure to UV light. Bands were scored using 1 to represent P_1 (N87-984-16), 2 to represent heterozygote, and 3 to represent P_2 (TN93-99) alleles for each primer locus.

Data analysis. The storage proteins (phenotypic data) β-conglycinin and glycinin, and their subunits α′, α, β, acidic and basic subunits from 2001 and 2003 were analyzed separately as well as combined using SAS software (23). Association between molecular marker and phenotype was determined by single-factor ANOVA utilizing PROC GLM of SAS using least square means from 2001, 2003, and combined data. With this analysis, molecular markers linked to QTL for storage protein and its components were identified. We confirmed the QTL by composite interval mapping (CIM) of QTL Cartographer version 2.0 using model 6 (standard model) with 10 cM window size, 2 cM walking speed, and five control molecular markers.

RESULTS AND DISCUSSION

The β-conglycinin (7S) content was derived by summation of the original scanned value of Alpha prime (α′), Alpha (α), and Beta (β), and the glycinin (11S) content was derived by summation of the acidic and basic components (Fig. 1). The values obtained from the densitometer scans were converted to percentage of the total protein in the lane. When selecting a genotype for higher glycinin content, it is conventional to calculate the ratio of 11S:7S (glycinin to β-conglycinin). This ratio was also compared across genotypes. Descriptive statistics for these components are presented in Table 1.

With respect to 7S, in developing a line with high S-containing essential amino acids, it is desirable to have the β subunit as low as possible since that subunit does not contain methionine or cysteine (24). The population in this study is promising in that regard, because we identified about 5% of the total lines with less than 6% β subunit (Table 1). The α′

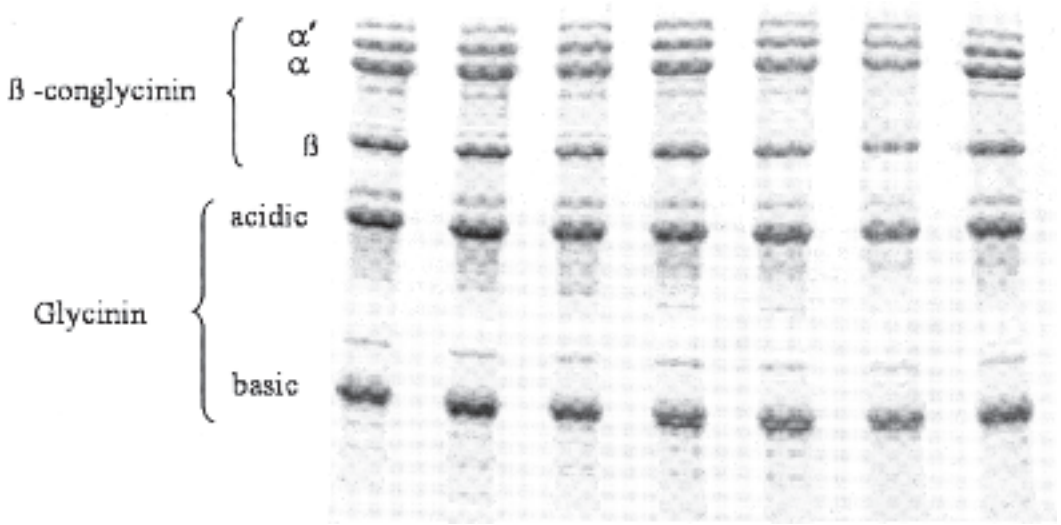

FIG. 1. Banding pattern of the components of 7S (α′, α, and β) and 11S (acidic and basic) fraction of soy protein in SDS-PAGE obtained from soybean recombinant inbred lines (RIL) of N87-984-16 × TN93-99.

TABLE 1
Descriptive Statistics of Soy Protein Fraction from a Population of 101 F_6-Derived RIL of N87-984-16 × TN93-99 (average of 2001 and 2003)

Traits	Minimum	Maximum	Mean	SD	P_1	P_2	Heritability (%)
α′	4.1	7.2	5.6	0.6	6.0	4.9	15.0
α	7.4	11.9	9.5	0.9	10.5	8.1	21.4
β	5.6	11.6	8.6	1.1	9.3	6.9	37.9
Conglycinin[a] (7S)	18.8	28.8	23.7	2.2	25.9	19.9	29.8
Acidic	16.8	23.3	19.6	1.2	21.0	16.8	29.3
Basic	12.5	17.9	15.5	1.0	16.4	13.2	0.0
Glycinin[a] (11S)	30.0	39.9	35.1	1.9	37.5	30.0	0.0
Ratio (11S:7S)	1.1	1.8	1.5	0.1	1.5	1.5	46.9
Sum (11S + 7S)	49.8	67.0	58.8	3.5	63.3	49.8	2.1

[a]Protein components reported are percent of total protein. RIL, recombinant inbred lines; P_1 = N87-984-16; P_2 = TN93-99.

and α subunits of 7S do contain traces of methionine and cysteine. Breeders targeting increased S-containing amino acids could select for those components, or develop a selection index that includes their sum. The sum of α′ and α ranged from 11 to 20% with a mean of 15%. This wide range gives us an opportunity to select lines from the upper range. Therefore, selection for a low level of the β subunit and high levels of α′ subunits of 7S could help increase the total S-containing amino acids.

Unlike 7S, it is not clear which subunit of 11S contains more S-containing amino acids. Both acidic and basic subunits showed a wide variation among lines. A moderate positive correlation ($r = 0.46$; $P < 0.01$) was found between acidic and basic subunits (Table 2). This indicates that both subunits increase in seeds simultaneously. Thus, selection for high levels of both the acidic and the basic subunits may increase the glycinin level. The level of variation in acidic and basic subunits is reflected in the high level of variation found in glycinin (Table 1). The overall distribution of 11S and 7S among the RIL is presented in Figures 2 and 3, respectively.

A high level of variation for 7S and 11S exists in this population. The scatter plot of 11S vs. 7S shows that glycinin content was high in the present population irrespective of β-conglycinin content (Fig. 4). This indicates that selection for lines with increased 11S could occur among a broad range of 7S levels, thereby providing an opportunity to ensure high-protein meal development. Indeed, Prolina appears to contribute to breaking the typical negative correlation between 7S and 11S, because such a correlation was not evident in our population (Fig. 4).

Since 11S contains significantly more S-containing amino acids per unit protein than does 7S, an increase in the 11S:7S ratio should lead to improvement in S-containing amino acids (15). Generally, the 11S:7S ratio is less than 1.5 in cultivated soybean, but it may be as high as 3.8 in wild *G. soja* ancestors of soybean (14,15). The existence of lines with a 11S:7S ratio of 1.8 in the present population (Table 1) clearly indicated that we developed RIL having more than typical 11S content. This is likely due to the N87-984-16 (Prolina parent), because Prolina was reported to have higher S-containing peptides than normal cultivars (25).

Yagasaki *et al.* (3) used soybean breeding lines having different glycinin subunit composition to determine 11S:7S ratios. The ratio varied from 0.04 to 1.48 among mutant genotypes lacking all three conglycinin subgroups. In a similar study So *et al.* (26) studied 58 cultivars of soybean according to their groupings (e.g., vegetable, sprout, sauce) and analyzed for different subunits of storage proteins. They determined the 11S:7S ratio in all cultivars and found that it ranged from 1.4 to 1.7. The cultivars with the highest 11S:7S ratio were in the vegetable type of cultivars. The ratio found in our study was greater than the highest value found by both these groups (3,26). This indicates that our population has more potential to develop a cultivar with high S-containing amino acids. Fehr *et al.* (27) analyzed 14 soybean cultivars for various storage proteins for 3 yr at eight locations using different

TABLE 2
Correlation Coefficients Between Conglycinin, Glycinin, and Their Ratio and Individual Components on the Basis of Combined Least Square Means from 2001 and 2003

	α′	α	β	Acidic	Basic	Conglycinin	Glycinin	11S:7S
α	0.63***							
β	0.56***	0.42***						
Acidic	0.61***	0.55***	0.33***					
Basic	0.01[NS]	0.14[NS]	0.06[NS]	0.46***				
Conglycinin	0.83***	0.80***	0.85***	0.56***	0.09[NS]			
Glycinin	0.40***	0.43***	0.24**	0.88***	0.82***	0.41***		
11S:7S	–0.64***	–0.59***	–0.76***	–0.04[NS]	0.41***	–0.81***	0.19[NS]	
11S + 7S	0.74***	0.74***	0.67***	0.85***	0.52***	0.86***	0.82***	–0.41***

[a]NS = Nonsignificant, * = significant at 0.05 probability, ** = significant at 0.01 probability, and *** = significant at 0.001 probability.

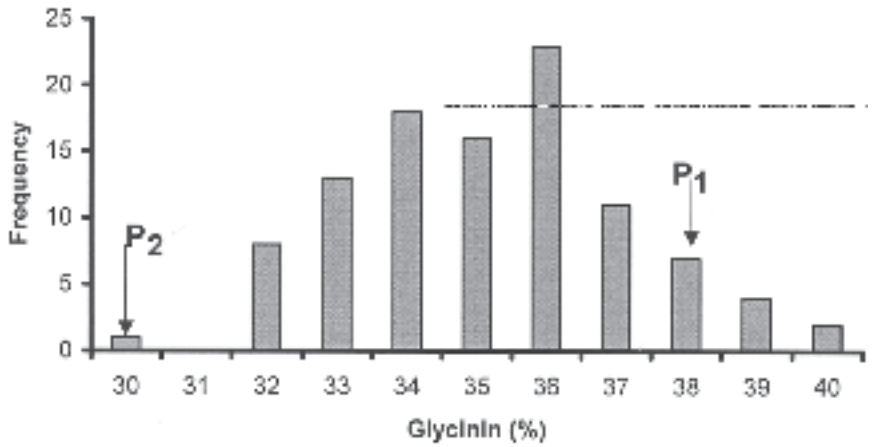

FIG. 2. Frequency distribution of glycinin in a population of F_6-derived RIL of N87-984-16 × TN93-99, where P_1 is N87-984-16 and P_2 is TN93-99. For other abbreviations see Figure 1.

groups of cultivars, including three with high-protein concentration. They found an 11S:7S ratio of 2.04 only in one high-protein cultivar; all others had a lower ratio. This suggests that high protein cultivars *per se* may not have high glycinin, and a need exists to utilize molecular markers (such as the ones we report in this paper) to identify the gene combinations that may enable improvements in S-containing amino acids.

We performed single-factor analysis to identify the candidate QTL for various components of storage protein. Then we performed CIM to confirm the QTL underlying storage protein and its subunits. We used CIM standard model (model 6) with a 10 cM window, 2 cM walking speed, and five control markers for the detection of QTL. The QTL for various components and their likelihood of odds (LOD) score are presented in Table 3.

QTL were detected that showed association between storage protein fractions and their subunits. Other QTL were detected that are associated with only one component. The QTL Satt461, located on linkage group (LG) D2, was associated with the α sub-unit of conglycinin in 2001 crop year samples. However, it was not detectable for the α subunit in 2003 samples as in the combined analysis. Therefore, it was environmentally specific and not a stable QTL (28). However, the same QTL (Satt461) was associated with the β subunit in 2003 samples and for total conglycinin of storage protein in the combined analysis (Table 3). Since α, α′, and β are the subunits of conglycinin, it is evident that the QTL Satt461 located on LG D2 is associated with the conglycinin gene family. Another QTL (Satt249) located on LG J was detected during the second year of the study and was associated with conglycinin and its β subunit (Table 3). In the combined analysis, Satt177, located on LG A2, was associated with the α′ subunit of conglycinin and Satt255, located on LG N, was associated with the β subunit. Therefore, it appears that more than one chromosomal region may govern the biosynthesis of conglycinin and its components.

The glycinin of the storage protein is composed of two subunits: acidic and basic. There were four QTL associated with the acidic and two QTL associated with the basic subunit of the glycinin fraction of storage protein in the combined data. QTL associated with the acidic subunit were Satt461, Satt518, Satt196, and Satt076, located on LG D2, K,

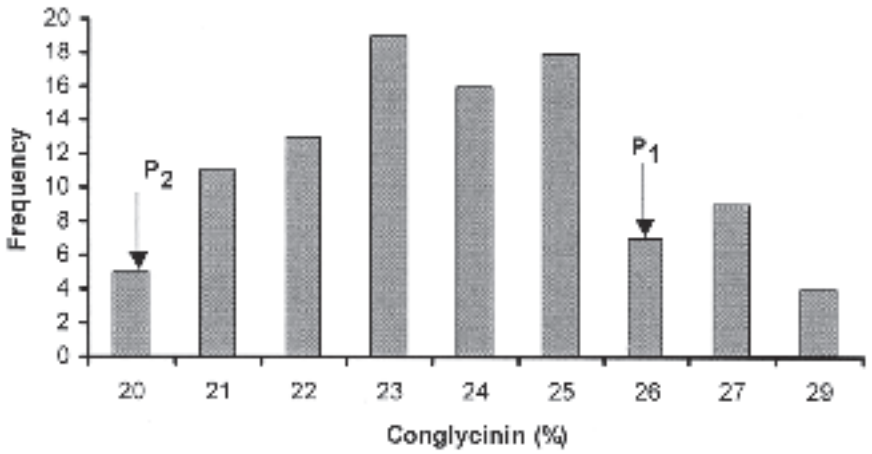

FIG. 3. Frequency distribution of β-conglycinin in a population of F_6- derived RIL of N87-984-16 × TN93-99. For other abbreviations see Figure 1.

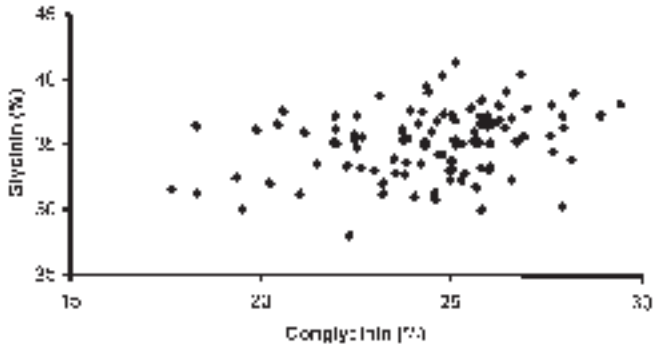

FIG. 4. Scatter plot of glycinin vs. β-conglycinin in RIL population developed from N87-984-16 × TN93-99. A high level of glycinin is present irrespective of β-conglycinin content. For other abbreviations see Figure 1.

K, and L, respectively. Since Satt461 was significant over years and in the combined data, it is a stable QTL associated with the acidic subunit of glycinin. The other QTL should be regarded as environmentally sensitive in their control of the acidic subunit. QTL associated with the basic subunit of glycinin were Satt292 and Satt156, located on LG I and L, respectively. Both of these QTL were environmentally stable in determining the concentration of the basic subunit of glycinin.

Six QTL were associated with the glycinin fraction of storage protein (Table 3). The most noteworthy were Satt461 (LG D2), which was significant for glycinin in the 2-yr combined analysis and Satt292 (LG I), which was significant for the basic subunit of the glycinin in the 2-yr combined analysis. Satt156 appeared to be environmentally sensitive because it was detected only during the second year for both the basic sub-unit and for glycinin. Two environmentally sensitive QTL (Satt191 and Satt564), both located on LG G, were detected only in the first year. The LOD score plot suggests that it may be useful to identify more polymorphic markers between the genomic regions of Satt564 and Satt191 to map the QTL and the positions controlling glycinin concentration more finely (Fig. 5).

TABLE 3
QTL Detected for Soybean Storage Protein Fractions by Composite Interval Mapping Analysis in F_6-Derived Population of N87-984-16 × TN93-99

Trait	Location	LG[a]	Marker	QTL position (cM)	LOD[a] score	Additive effect[b]	R^2
α	2001	D2	Satt461	93.4	2.1	–0.4	13.4
α′	2001	K	Satt196	50.3	4.3	3.2	16.7
	2001	K	Satt102	57.3	4.7	4.1	27.0
	2001	L	Satt076	8.0	2.9	–3.3	15.5
	2001	M	Satt540	35.8	3.1	–3.7	19.9
	Combined	A2	Satt177	113.5	2.0	–2.5	14.9
β	2003	D2	Satt461	99.4	2.7	–0.6	12.4
	2003	J	Satt249	64.5	2.2	0.6	10.1
	Combined	N	Satt255	0.0	2.3	–0.4	9.1
Conglycinin	2003	J	Satt249	64.5	2.3	1.0	10.3
	Combined	D2	Satt461	99.4	2.1	–0.7	11.6
Acidic	2001	D2	Satt461	71.4	2.8	–0.9	36.6
	2003	D2	Satt461	99.4	2.3	–0.6	11.2
	Combined	D2	Satt461	99.4	3.3	–0.5	14.0
	Combined	K	Satt518	28.0	2.1	0.4	8.1
	Combined	K	Satt196	32.3	2.2	0.4	10.6
	Combined	L	Satt076	10.0	2.5	–0.4	11.7
Basic	2001	A2	Satt133	76.6	2.3	0.6	12.7
	2001	I	Satt292	0.0	3.0	0.6	13.7
	2003	A2	Satt177	117.5	2.6	–0.5	15.9
	2003	L	Satt156	100.5	3.3	0.6	17.6
	Combined	I	Satt292	0.0	2.6	0.4	11.9
	Combined	L	Satt156	104.5	2.1	0.3	10.3
Glycinin	2001	A2	Satt437	65.3	2.6	1.0	12.8
	2001	G	Satt564	14.0	3.7	–1.3	21.8
	2001	G	Satt191	23.2	3.3	–1.2	18.6
	2001	I	Satt292	0.0	2.0	0.8	9.5
	2003	D2	Satt461	99.4	2.3	–1.0	11.2
	2003	L	Satt156	100.5	2.5	1.2	16.0
	Combined	D2	Satt461	87.4	2.3	–0.8	19.5

[a]LG, linkage group; LOD, likelihood of odds.
[b]Additive quantitative trait loci (QTL) effect is with respect to P_1 (N87-984-16).

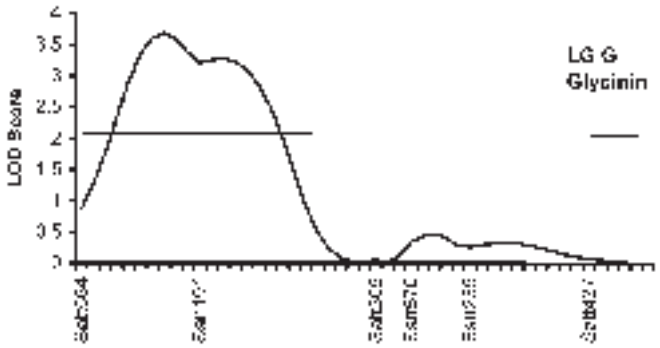

FIG. 5. Likelihood of odds (LOD) score plot for glycinin on linkage group (LG) G. A major quantitative trace locus (QTL) appears located between Satt564 and Satt191, and another major QTL appears located downstream of Satt191.

Detection of QTL with individual components of conglycinin and glycinin but nonsignificant association with glycinin and conglycinin suggests an interaction of genes governing folding and unfolding of the individual subunit of the fractions of the storage protein. The present experiment identified major QTL for conglycinin, glycinin, and their components. These QTL may be useful in conducting marker-assisted selection in soybean breeding programs.

It is believed that 11S is rich in S-containing amino acid; hence, it is desirable to have greater concentrations of 11S. In this experiment, we found one common QTL (Satt461) for 11S and 7S. From an MAS point of view, a negative correlation among protein fractions is not a favorable condition because we want to select only for 11S, not for 7S. Therefore, the common QTL (Satt461) cannot be very useful to serve as a unique marker for this purpose. However, there are other QTL (Satt292 and Satt156) associated with 11S but not with 7S. These two QTL can be used as unique QTL in MAS to improve the nutritional quality of soybean. It should be noted that we attempted to be conservative, reporting only QTL that were significant based both on single-factor analysis and CIM (Table 3). The environmentally sensitive QTL reported here could be considered stable based on results from single-factor ANOVA. Indeed, Stuber *et al.* (29) have shown that single-factor analysis is as effective as interval mapping, and most soybean protein QTL reported to date have been detected using single-factor analysis (SOYBASE, http://www.soybase.org). We anticipate that the QTL reported in this paper will enhance opportunities to improve protein quality in soybean.

A high ratio of 11S:7S may increase the S-containing essential amino acids, hence, identification of markers associated with QTL governing 11S and 7S is important to geneticists to identify individuals that are likely to enable genetic gains for protein quality. Parental allelic contributions for a trait were expressed with respect to P_1 (N87-984-16) as an additive effect (Table 3). Two markers (Satt156 on LG L and Satt292 on LG I) associated with QTL for glycinin were associated with a positive contribution from P_1, whereas Satt461 (LG D2) was contributed from P_2 (TN93-99). This indicates that desirable allele frequencies can be increased at those loci to augment genetic gains for glycinin. Therefore, the desirable alleles would need to be selected from both the parents, and selected individuals could be identified that inherit the best combination of alleles, leading to glycinin levels that exceed that of the higher parent (Fig. 2). The increase of desirable allele frequencies from both parents in individual RIL is a major goal of modern molecular breeding programs. For example, Qiu *et al.* (30) conducted an experiment and found alleles from both the parents were beneficial for increasing the protein content in a population derived from 'Peking' × 'Essex' soybean.

An interesting aspect of our experiment was that Satt461 was significantly associated with conglycinin as well as glycinin fractions of seed storage protein. This is not a particularly desirable condition for an MAS program, because we may want to have reduced 7S and increased 11S content for developing a high S-containing amino acid genotype. If the same marker is associated with both traits, presumably there is pleiotropy or tight genetic linkage. Alternatively, marker Satt292 and Satt156 (located on LG I and L) were found to have a positive contribution for glycinin, and they mapped to a linkage group with no other markers associated with conglycinin. Brummer *et al.* (28) found two QTL, located on LG I and L, for protein concentration in soybean but they did not analyze for glycinin. Utilization of our glycinin markers for an MAS in developing a line with increased 11S:7S ratio may be possible.

Nielsen *et al.* (9) identified five genes (*Gy1* through *Gy5*) regulating the production of glycinin in soybean. They grouped the five genes into two groups: three genes in group I and two genes in group II. Nielsen *et al.* (9) and Cho *et al.* (31) reported that all five genes assort independently, indicating five different chromosomes. In our experiment, we found molecular markers distributed on three different LG (D2, I, and L) conferring glycinin concentration (Table 3). Moreover, Beilinson *et al.* (32) found two additional genes (*Gy6* and *Gy7*) that mapped on LG N and L, respectively. This indicates that we may not have mapped all QTL in this population; thus further research is warranted. To detect more protein-quality QTL, development of additional diverse populations would be useful.

Harada *et al.* (10) identified three genes that control the production of β-conglycinin. Furthermore, Nielsen (33) reported that genes encoding for α and β are linked, whereas α′ and α assort independently, indicating that there should be at least two linkage groups containing the genes encoding β-conglycinin. We found two QTL located on two different LG (D2 and J) governing β-conglycinin (Table 3).

Increased glycinin content in soybean protein is an important trait for increasing the concentration of the S-containing amino acids. The population developed for this study has been used successfully with a molecular breeding approach, resulting in the recent release of soybean germplasm TN04-5363 by the Tennessee Agricultural Experiment Station. This germplasm expresses significantly higher levels of cysteine, and total sulfur-containing amino acid, than P_1 and P_2. It is being made available to breeders to improve soybean protein

quality. The QTL reported in this paper provide new information for soybean geneticists for targeting the development of soybean lines with enhanced protein quality.

ACKNOWLEDGMENTS

Funding from the United Soybean Board, Better Bean Initiative, Project 1240 and by the Tennessee Soybean Promotion Board is greatly appreciated. Continued support by the Tennessee Agriculture Experiment Station is also highly appreciated. We wish to express our thanks to Dr. Madeleine Spencer and to Deborah Landau-Ellis for their contribution to this project.

REFERENCES

1. Clarke, E.J., and J. Wiseman, Developments in Plant Breeding for Improved Nutritional Quality of Soybeans I. Protein and Amino Acids Content, *J. Agric. Sci. 134*:111–124 (2000).
2. Tsukada, Y., K. Kitamura, K. Harada, and N. Kaizuma, Genetic Analysis of Two Major Sub-units (β-conglycinin and glycinin) in Soybean Seeds, *Jpn. J. Soybean 36*:390–400 (1986).
3. Yagasaki, K., T. Takagi, M. Sakai, and K. Kitamura, Biochemical Characterization of Soybean Protein Consisting of Different Subunits of Glycinin, *J. Agric. Food Chem. 45*:656–660 (1997).
4. Kitamura, K., Genetic Improvement of Nutrition and Food Processing Quality in Soybean, *Jpn. Agric. Res. Quart 29*:1–8 (1995).
5. Ogawa, T., E. Tayama, K. Kitamura, and N. Kaizama, Genetic Improvement of Seed Storage Proteins Using Three Variant Alleles of 7S Globulin Sub-Unit in Soybean (*Glycine max* L.), *Jpn. J. Breeding 39*:137–147 (1989).
6. Kitagawa, S., M. Ishimoto, F. Kikuchi, and K. Kitamura, A Characteristic Lacking or Decreasing Remarkably 7S Globulin Sub-unit Induced with X-Ray Irradiation in Soybean Seeds, *Ibid. 41*:460–461 (1991).
7. Takahashi, K., Y. Mizumo, S. Yumoto, K. Kitamura, and S. Nakamura, Inheritance of the α-Sub-unit Deficiency of β-Conglycinin in Soybean (*Glycine max* L.) Line Induced by γ-Ray Irradiation, *Breeding Sci. 46*:251–255 (1996).
8. Yaklich, R.W., β-Conglycinin and Glycinin in High-Protein Soybean Seeds, *J. Agric. Food Chem. 49*:729–735 (2001).
9. Nielsen, N.C., C.D. Dickinson, T.J. Cho, V.H. Thanh, B.J. Scallon, R.L. Fischer, T.L. Sims, G.N. Drews, and R.B. Goldberg, Characterization of the Glycinin Gene Family in Soybean, *Plant Cell 1*:313–328 (1989).
10. Harada, J.J., S.J. Barker, and R.B. Goldberg, Soybean β-Conglycinin Genes Are Clustered in Several DNA Regions and Are Regulated by Transcriptional and Posttranscriptional Processes, *Ibid. 1*:415–425 (1989).
11. Hayashi, M., M. Nishioka, K. Kitamura, and K. Harada, Identification of AFLP Markers Tightly Linked to the Gene for Deficiency of the 7S Globulin in Soybean Seed and Characterization of Abnormal Phenotypes Involved in the Mutation, *Breeding Sci.50*:123–129 (2000).
12. Kitamura, K., and N. Kaizuma, Mutant Strains with Low Level of Subunits of 7S Globulin in Soybean (*Glycine max* Merr.) Seed, *Jpn. J. Breeding 31*:353–359 (1981).
13. Mandal, S., and R. Mandal, Seed Storage Proteins and Approaches for Improvement of Their Nutritional Quality by Genetic Engineering, *Curr Sci. 79*:576–589 (2000).
14. Kwanyuen, P., V.R. Pantalone, J.W. Burton, and R.F. Wilson, A New Approach to Genetic Alteration of Soybean Protein Composition and Quality, *J. Am. Oil Chem. Soc. 74*:983–987 (1997).
15. Liu, K., *Soybeans: Chemistry, Technology and Utilization*, Chapman and Hall, London, 1997, pp. 478–523.
16. Burton, J.W., T.E. Carter, and R.F. Wilson, Registration of Prolina Soybean, *Crop Sci. 39*:294–295 (1999).
17. Luck, P., T. Lanier, C. Daubert, R.F. Wilson, and P. Kwanyuen, Functionality and Visoelastic Behavior of Prolina Soybean Isolate, in *Oilseed Processing and Utilization*, edited by R.F. Wilson, AOCS Press, Champaign, 2001, pp. 197–202.
18. Pantalone, V.R., F.L. Allen, and D. Landau-Ellis, Registration of 'Tn93-99' Soybean Germplasm, *Crop Sci. 43*:1137 (2003).
19. Bradford, M.M., A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding, *Anal. Biochem. 2*:248–254 (1976).
20. Chua, N.H., Electrophoretic Analysis of Chloroplast Proteins, *Methods Enzymol. 69*:434–446 (1980).
21. Kwanyuen, P., and R.F. Wilson, Optimization of Coomassie Staining for Quantitative Densitometry of Soybean Storage Proteins in Gradient Gel Electrophoresis, *J. Am. Oil Chem. Soc. 77*:1251–1254 (2000).
22. Cregan, P.B., T. Jarvik, A.L. Bush, R.C. Shoemaker, K.G. Lark, A.L. Kahler, N. Kaya, T.T. VanToai, D.G. Lohnes, and J. Chung, An Integrated Genetic Linkage Map of the Soybean Genome, *Crop Sci. 39*:1464–1490 (1999).
23. S*AS Guide for Personal Computers*, SAS Institute, Cary, NC, 2000.
24. Krishnan, H.B., Biochemistry and Molecular Biology of Soybean Seed Storage Proteins, *J. New Seeds 2*:1–25 (2000).
25. Wilson, R.F., Current Research on Oilseeds, in *Institute of Food Technologists 1996 Symposium: Identity-Preserved Oils*, symposium held at the annual meeting of the Institute of Food Technologists, New Orleans, 1996.
26. So, E., Y. Chae, Y. Kim, and Y. MH, Variation of 7S and 11S Seed Protein Concentrations in Different Food Types of Soybean Seed, *Korean J. Crop Sci. 44*:350–354 (1999).
27. Fehr, W.R., J.A. Hoeck, S.L. Johnson, P.A. Murphy, J.D. Nott, G.I. Padilla, and G.A. Welke, Genotype and Environment Influence on Protein Components of Soybean, *Crop Sci. 43*:511–514 (2003).
28. Brummer, E.C., G.L. Graef, J. Orf, J.R. Wilcox, and R.C. Shoemaker, Mapping QTL for Seed Protein and Oil Content in Eight Soybean Populations, *Ibid. 37*:370–378 (1997).
29. Stuber, C., S. Lincoln, D. Wolff, T. Helentjaris, and E. Lander, Identification of Genetic Factors Contributing to Heterosis in a Hybrid from Two Elite Maize Inbred Lines Using Molecular Markers, *Genetics 132*:823–839 (1992).
30. Qiu, B.X., P.R. Arelli, and D.A. Sleper, RFLP Markers Associated with Soybean Cyst Nematode Resistance and Seed Composition in a 'Peking' × 'Essex' Population, *Theor. Appl. Genet. 98*:356–364 (1999).
31. Cho, T.J., C.S. Davies, and N.C. Nielsen, Inheritance and Organization of Glycinin Genes in Soybean, *Plant Cell 1*:329–337 (1989).
32. Beilinson, V., Z. Chen, R.C. Shoemaker, R.L. Fischer, R.B. Goldberg, and N.C. Nielsen, Genomic Organization of Glycinin Genes in Soybean, *Theor. Appl. Genet. 104*:1132–1140 (2002).
33. Nielsen, N., Soybean Seed Composition, in *Soybean: Genetics, Molecular Biology and Biotechnology*, edited by D. Verma, and R. Shoemaker, CAB International, Wallingford, 1996, pp. 127–163.

[Received May 11, 2004; accepted October 21, 2004]