Theor Appl Genet (1996) 93:649–657

S. H. Lee · M. A. Bailey · M. A. R. Mian
T. E. Carter, Jr · E. R. Shipe · D. A. Ashley
W. A. Parrott · R. S. Hussey · H. R. Boerma

RFLP loci associated with soybean seed protein and oil content across populations and locations

Received: 14 January 1996 / Accepted: 8 March 1996

Abstract Molecular markers provide the opportunity to identify marker-quantitative trait locus (QTL) associations in different environments and populations. Two soybean [*Glycine max* (L.) Merr.] populations, 'Young' × PI 416 937 and PI 97 100 x 'Coker 237', were evaluated with restriction fragment length polymorphism (RFLP) markers to identify additional QTLs related to seed protein and oil. For the Young × PI 416 937 population, 120 F_4-derived lines were secored for segregation at 155 RFLP loci. The F_4-derived lines and two parents were grown at Plains, G.a., and Windblow and Plymouth, N.C. in 1994, and evaluated for seed protein and oil. For the PI 97 100 × Coker 237 population, 111 F_2-derived lines were evaluated for segregation at 153 RFLP loci. Phenotypic data for seed protein and oil were obtained in two different locations (Athens, G.a., and Blackville, S.C.) in 1994. Based on single-factor analysis of variance (ANOVA) for the Young × PI 416937 population, five of seven independent markers associated with seed protein, and all four independent markers associated with seed oil in the combined analysis over locations were detected at all three locations. For the PI 97 100 × Coker 237 population, both single-factor ANOVA and interval mapping were used to detect QTLs. Using single-factor ANOVA, three of four independent markers for seed protein and two of three independent markers for seed oil were detected at both locations. In both populations, single-factor ANOVA, revealed the consistency of QTLs across locations, which might be due to the high heritability and the relatively few QTLs with large effects conditioning these traits. However, interval mapping of the PI 97 100 × Coker 237 population indicated that QTLs identified at Athens for seed protein and oil were different from those at Blackville. This might result from the power of QTL mapping being dependent on the level of saturation of the genetic map. Increased seed protein was associated with decreased seed oil in the PI 97 100 × Coker 237 population ($r = -0.61$). There were various common markers ($P \leqslant 0.05$) on linkage groups (LG) E, G, H, K, and UNK2 identified for both seed protein and oil. One QTL on LG E was associated with seed protein in both populations. The other QTLs for protein and oil were population specific.

Key words Soybean *Glycine max* · Protein content Oil content · Mapping · QTL · RFLP

Communicated by J. Mac Key

S. H. Lee · M. A. Bailey[1] · M. A. R. Mian · D. A. Ashley · W. A. Parrott · H. R. Boerma (✉)
Department of Crop and Soil Sciences, University of Georgia, Athens, GA 30602-7272, USA

T. E. Carter, Jr
USDA-ARS Dep. of Crop Science, North Carolina State University, Raleigh, NC 27695-7631, USA

E. R. Shipe
Department of Agronomy, Clemson University, Clemson, SC 29634-0359, USA

R. S. Husseey
Department of Plant Pathology, University of Georgia, Athens, GA 30602-7274, USA

Present address:
[1]Pioneer Hi-Bred Intl., 7300NW 62nd Ave., P.O.Box 1004, Johnston, IA 50131, USA

Introduction

Because soybean seed is a major source of protein for animal feed and oil for human consumption, breeding efforts have been directed toward the improvement of seed protein and oil content. These efforts have resulted in the selection of two types of cultivars; those with a higher percentage of protein in the seed and those with a higher oil content (Fehr and Weber 1968; Smith and Weber 1968; Brim and Burton 1979; Burton and Brim 1981; Burton 1985; Wilcox 1985). The substantial progress in this area was possible due to moderately high heritabilities of these traits. Broad-sense heritability estimates for the percentage of protein range for 0.57 to 0.91, and those for the percentage of oil from 0.51 to 0.93

(Fehr and Weber 1968; Smith and Weber 1968; Kwon and Torrie 1964; Byth et al. 1969).

With the advent of molecular markers, QTL mapping is becoming an important tool to facilitate the selection of quantitative characters in crop-breeding programs (Tanksley et al. 1989). Molecular markers have been used widely for the detection and location of QTLs underlying quantitative variation in the genome (Lander and Botstein 1989), the development and improvement of inbred lines (Dudley 1994), and the determination of the genetic basis of heterosis (Stuber et al. 1992; Xiao et al. 1995).

In soybean, a limited number of studies have identified QTLs associated with agronomically important traits on the two initial public RFLP maps (Keim et al. 1990b; Lark et al. 1993). Keim et al. (1990b) detected QTLs associated with leaf width and length, stem diameter and length, canopy height, and maturity in 60 F_2-derived lines from a cross between *G. max* and *G. soja*. Additionally, in the same population, five independent genomic regions were identified as associated with seed hardness (Keim et al. 1990a). Using interval mapping, Mansur et al. (1993) reported that 11 of 15 reproductive, morphological and seed traits were localized to genomic intervals in six different linkage groups.

To identify QTLs associated with soyban seed protein and oil, Diers et al. (1992) evaluated an F_2 population derived from a cross between *G. max* and *G. soja*. They found three independent markers associated with seed protein located on linkage groups (LGs) E, F, and I (Shoemaker and Specht 1995). For seed oil, two genomic regions (LG E and I), consistent with QTLs for protein, were identified. In an F_2-derived soybean population from a 'Minsoy' × 'Noir 1' cross, Mansur et al. (1993) reported an unlinked RFLP locus, L48, associated with seed protein. Two genomic regions, near T153a (LG A2, Shoemaker and Specht 1995) and A315 (possibly LG K), were associated with seed oil. Comparison of these two studies on the position of QTLs associated with seed protein and oil indicated the different genomic location of putative QTLs from different populations and provided support for the population specificity of important QTLs for oil and protein content.

Most previous studies that examined the association of molecular markers with QTLs in soybean were evaluated in a single environment with a single population. In our previous study of QTLs associated with plant height, lodging, and maturity in the 'Young' × PI 416937 soybean population, grown in four different locations, the level of consistency of QTLs across locations was trait specific (Lee et al. 1996a). The QTLs for plant height and lodging were clearly inconsistent, whereas those for maturity showed good agreement across locations. A consistency of QTLs across locations is supported by Stuber et al. (1992). In contrast, there have been numerous reports indicating the inconsistency of QTLs across environments (Paterson et al. 1991; Bubeck et al. 1993; Dudley 1993). In another soybean population, PI 97100 × 'Coker 237', which segregates for growth habit (Lee et al. 1996b), the QTLs associated with plant height, lodging, and maturity were consistent across two locations, which probably resulted from a few loci with large effects conditioning the metric traits. This suggested that the level of consistency of QTL association was dependent on the types of traits evaluated, as well as the genetic background of the populations.

Until now, studies of QTLs associated with soybean seed protein and oil have collected data from only two populations and a limited sample of environments (Diers et al. 1992; Mansur et al. 1993). Diers et al. (1992) identified QTLs for seed protein and oil on the basis of data averaged across three environments, but did not report on the consistency of QTLs across environments.

The objective of the present study was to identify additional QTLs associated with seed protein and oil in two soybean populations. A secondary purpose was to examine the consistency of QTLs across several locations and the population specificity of QTLs.

Materials and methods

Two soybean populations were studied to construct genetic linkage maps, and were evaluated for seed protein and oil content. Population I was F_4-derived from a cross of the productive cultivar Young (Burton et al. 1987) and the drought-tolerant plant introduction PI 416937 (Sloane et al. 1990). This population consisted of 120 lines which were created by single-seed-descent, with each line originating from a different F_2 plant. Population II, PI 97100 × Coker 237, was segregating for growth habit. PI 97100 posseses an indeterminate growth habit, and Coker 237 has a determinate growth habit. From this second population, a total of 111 F_2-derived lines were developed, with each line originating from a different F_2 plant.

RFLP genotypes of each line of the two populations had been determined previously (Lee et al. 1996a, b) using probes from various sources, including cDNA and/or genomic clones of soybean (R. C. Shoemaker, USDA. Iowa State Univ.; K. G. Lark, Univ. of Utah; R. T. Nagao, Univ. of Georgia), *Vigna radiata* (N. D. Young, Univ. of Minnesota), *Phaseolus vulgaris* (J. M. Tohme, CIAT), *Arachis hypogaea* (G.D. Kochert, Univ. of Georgia), and *Medicago sativa* (G. D. Kochert). The DNA isolation, Southern blotting, and hybridization procedures have been described previously (Lee et al. 1996a,b). The linkage map was constructed with marker genotypic data using the Kosambi map function of GMendel (Holloway and Knapp 1993) and Mapmaker (Lander et al. 1987). For the F_4-derived population of Young × PI 416937, the genetic map was reconstructed with GMendel (Mian et al. 1996) due to the expansion of genetic map distance that occurs in advanced lines (Haldane and Waddinton 1931; Burr et al. 1988). For the F_2-derived Coker 237 × PI 97100 population, the genetic map was constructed with Mapmaker (Lee et al. 1996b). For grouping markers into linkage groups, a minimum LOD of 3.0 and a maximum distance of 50 cM (rmax of 0.38 for GMendel) were used.

For phenotypic data, the parents and 120 lines from Young × PI 416937 were grown in 1994 at Plains (Univ. of Georgia Southwest Branch Exp. Stn.), G. a., and Windblow (North Carolina State Univ. Sandhills Res. Stn.) and Plymouth (North Carolina State Univ. Plymouth Res. Stn.), N.C. The parents and 111 lines from PI 97100 × Coker 237 were grown in 1994 at two locations, Athens (Univ. of Georgia Plant Sciences Farm), G. a., and Blackville (Clemson Univ. Edisto Research and Educational Centre), S.C. To reduce experimental error due to soil heterogeneity within each experimental site, the lines in each population were divided into three groups. The lines in each of these groups were placed in three separate tests, along with three entries of Young and one entry of PI 416937 as reference genotypes for the Young × PI 416937 population, and with PI 97100,

Coker 237, and 'Stonewall' for the PI 97 100 × Coker 237 population. Each test for Population I was grown in two replications of a randomized complete block design, except at Windblow, N. C., which had three replications. In Population II, two tests were grown in four replications, and one test in three replications (due to seed availability) of a randomized complete block design. Plot sizes for each location were previously described (Lee et al. 1996a,b).

Protein and oil percentages were determined on a plot basis. Seed was sent to USDA-ARS, National Center for Agricutural Utilization Research at Peoria, Ill. for analysis. Two samples per plot of 18–20 g of seed were analyzed for protein and oil composition with a Infratec NIR Food and Feed Grain Analyzer (Model 1255). Analysis of the seed was conducted on an as is basis and then mathematically converted to a moisture-free basis.

The association between marker and QTL was tested using the interval mapping method (Lander and Botstein 1989) with Mapmaker-QTL software (Lincoln et al. 1992) and by single-factor ANOVA. QTL analyses were performed on the standardized mean data within each location as well as across locations (Lee et al. 1996a). A LOD score of 2.5 was chosen as a minimum to declare the presence of a QTL in a given genomic region. The LOD score peak was used to estimate the most likely QTL position on the RFLP linkage map. The percentage of variance explained by individual QTLs and the additive (a) and dominance (d) effects were estimated at the maximum likelihood QTL position. The average degree of dominance for each QTL was calculated as the ratio d/a. Single-factor ANOVA was also used to determine the significance among RFLP genotypic class means using an F-test from the Type-III mean squares obtained from the GLM Procedure (SAS 1988). In the F_4-derived Young × PI 416 937 population, only homozygous RFLP classes were selected at each marker locus, and were compared for the determination of significant difference. The probability level was $P \leqslant 0.01$ for the combined analysis over locations, and $P \leqslant 0.05$ for each individual location. The lower probability level in each individual location was chosen to enhance the ability to detect QTL consistency across locations. Two-way analysis of variance was used to detect epistatic interactions between markers with significant associations.

Results

Genetic map

A total of 155 RFLP markers for the Young × PI 416 937 population and 153 for the PI 97 100 × Coker 237 population were used to construct genetic linkage maps of soybean (Table 1). For the Young × PI 416 937 population, 140 markers were genetically linked and were placed into 33 linkage groups covering 970 cM. The map of the PI 97 100 × Coker 237 population covered about 1600 cM with 142 markers classified into 23 linkage groups.

Table 1 Number of progeny, RFLP loci, linkage groups, and estimated genome size for two soybean populations

Characteristic	Young × PI 416 937	PI 97 100 × Coker 237
Progeny (no.)	120	111
Generation	F4-derived	F2-derived
RFLP loci (no.)	155	153
Linked RFLP loci (no.)	140	142
Linkage groups (no.)	33	23
Estimated genome size (cM)	970	1600
Average interval between two markers (cM)	6.	11.3

The map of the Young × PI 416 937 population contains 47 heterologous probes, with 37 from *A. hypogaea*, nine from *M. sativa*, and one from *P. vulgaris*. A total of 36 heterologous probes with 26 from *A. hypogaea*, five from *M. sativa*, and five from *P. vulgaris* mapped in the PI 97 100 × Coker 237 population. These heterologous probes were distributed among most of linkage groups in both maps.

Young × PI 416 937

Seed protein

Young and PI 416 937 differed by only 11 g/kg in seed protein content. However, transgressive variation of 38 g/kg occurred among progenies (Table 2). Using single-factor ANOVA for the combined data over three locations, 13 RFLP loci were associated with seed protein (Table 3). Seven of these represent putative independent QTLs [greater than 50 cM from another RFLP locus significantly ($P \leqslant 0.05$) associated with the trait, and the RFLP locus acting in an additive manner with regard to explaining variation]. In all cases, when two putative independent QTLs were identified on the same linkage group, there was at least one RFLP locus between the putative independent QTL that was not significantly ($P \geqslant 0.05$) associated with seed protein. At all independent loci, except A517-1, the PI 416 937 allele increased seed protein.

Individually, the seven independent QTLs for protein accounted for 6.8 to 13.8% of the phenotypic variation. The amount of variation explained by these seven independent markers sums to 69.5%. The heritability (asuming a selection unit of three locations, two replications/location) for seed protein was 83.2% (Table 2). Thus, when combined, these seven markers explain most of the genetic variation for seed protein in this population. All possible pairwise combinations of the putative independent markers associated with seed protein were tested for two-factor interactions to detect epistasis, but none was detected for any of these markers.

Table 2 Means and ranges of parents and progeny for soybean seed protein and oil, and their heritabilities

Item	Young × PI 416 937		PI 97 100 × Coker 237	
	Protein (g/kg)	Oil (g/kg)	Protein (g/kg)	Oil (g/kg)
Female parent	441	202	451	186
Male parent	452	184	426	200
Progeny range	431–469	182–201	427–462	182–203
LSD (0.05)	8.0	3.6	10.7	7.2
Progeny ×	447	192	445	190
h^2(%)[a]	83.2	89.1	82.8	74.9

[a] Selection unit = three locations, two replications/location for the Young × PI 416 937 population, and two locations, three replications/location for the PI 97 100 × Coker 237 population

Table 3 RFLP loci associated with seed protein and oil in the F_4-derived soybean population of Young × PI 416937

Trait	RFLP locus	Linkage group[a]	R^2(%)	Genotypic means[b] (g/kg)		R^2(%)		
				A/A	B/B	Plain	Windblow	Plymouth
Protein	gac197-1[c]	C1	12.5	443	449	8.3	5.9	13.4
	EV3-1	C1	11.1	443	448	8.8	5.8	9.6
	A338-1n	C1	10.1	443	448	14.0	–	6.2
	A463-1[c]	C1	6.8	444	448	–	9.1	8.1
	A517-1[c]	E	7.4	448	443	4.9	6.9	4.6
	cr167-1	E	6.5	448	443	4.4	6.4	–
	B166-1[c]	J	7.6	444	449	–	5.0	8.8
	A071-1[c]	N	11.2	443	448	8.3	6.6	9.1
	gac34-2	N	7.6	444	448	–	6.5	8.5
	B142-1[c]	P	10.2	443	448	8.6	6.6	6.6
	A352-1	P	10.0	443	449	6.6	7.2	8.2
	A199-3[c]	UNK1	13.8	442	449	7.9	8.4	15.8
	cr274-2	UNK1	8.0	444	449	5.4	5.7	6.6
Oil	A069-3[c]	E	6.7	193	191	6.2	4.0	5.2
	B122-1[c]	J	7.0	194	191	4.0	5.4	6.5
	A023-1[c]	L	7.4	191	193	5.7	4.9	6.2
	cr142-1[c]	R	12.9	194	191	14.0	13.4	4.6
	K258-1	R	8.8	194	191	8.2	12.3	–
	cr326-1n	R	8.8	193	191	10.3	9.1	–

[a] Based on the designation of Shoemarker and Specht (1995)
[b] A/A: homozygous Young, B/B: homozygous PI 416937
[c] Putative independent QTL (greater than 50 cM from another marker significantly associated with seed protein and oil, and additive gene action with regards to explaining variation)

Of the seven independent markers associated with seed protein in the combined analysis, five were detected at all three locations. Thus, the RFLP markers for seed protein in this population were mostly consistent across the three locations. The genotype × location variance component was 26.0% of the genotypic variance component for seed protein.

Seed oil

Young and PI 416937 differed by 18 g/kg in seed oil content, and seed oil content varied from 182 to 201 g/kg among progenies (Table 2). For seed oil, six markers were detected on the basis of the combined analysis over three locations (Table 3). Four of these six markers represent putative independent QTLs associated with seed oil. At all of these four loci, except A023-1, the Young allele was associated with increased seed-oil content.

The cr142-1 marker on LG R explained 12.9% of the total variation for seed oil. Summed together, four independent markers explained 34.0% of the total variation for seed oil. The heritability (assuming a selection unit of three locations, two replications/location) for seed oil was 89.1% (Table 2). Therefore, most of the genetic variation in seed oil was not explained by these four independent markers, suggesting that undetected QTLs exist or that the identified RFLP markers are a considerable distance from the QTL for seed oil in this population. There were no epistatic interactions identified among the six two-way combinations betwen two independent markers.

All four independent markers were significant ($P \leqslant 0.05$) at all three locations, indicating the consistency of these QTLs for seed oil across locations. The genotype × location variance component for seed oil was 12.2% of the genotypic variance component.

PI 97100 × Coker 237

Seed protein

PI 97100 and Coker 237 differed by 25 g/kg in seed protein, and seed protein varied among F_2-derived lines from 427 to 462 g/kg (Table 2). Single-factor ANOVA revealed that six RFLP loci were detected on the basis of combined analysis over two locations (Table 4). On LG K, three markers (A065-1, R051-3, Q43-1) were found to be significant ($P \leqslant 0.01$). The N100-1 locus between A065-1 and R051-3 was also detected as significant ($0.01 < P < 0.05$). However, Q45-1 between A065-1 and Q43-1 was not significant in the combined analysis. The Q45-1 locus was associated with seed protein in Blackville ($P \leqslant 0.05$). The existence of nonsignificant markers between two markers associated with seed protein on LG K was probably caused by data missing for the Q45-1 and N100-1 loci, as RFLP data were obtained from only 95 and 89 of 111 lines, respectively. Therefore, these three markers on LG K were assumed to detect the same QTL, in spite of the existence of a nonsignificant marker between them. Three other markers, located on LG E, H, and an unknown linkage group (designated as UNK2), represent putative inde-

Trait	RFLP locus	Linkage group[a]	R^2(%)	Genotypic means[b] (g/kg)			R^2(%)	
				A/A	A/B	B/B	Athens	Blackville
Protein	A454-1[c]	E	8.8	449	445	442	7.4	7.1
	A566-2[c]	H	13.5	451	445	443	11.4	10.8
	A065-1[c]	K	10.6	442	444	449		18.7
	R051-3	K	10.2	443	445	450		12.0
	Q43-1	K	9.6	443	444	449	–	13.4
	A132-4[c]	UNK2	13.3	441	446	448	13.5	9.1
Oil	A063-1[c]	C1	13.2	191	191	188	11.4	7.3
	L154-2[c]	G	17.1	189	192	189	21.4	–
	A235-1	G	14.7	188	192	189	16.5	8.0
	L002-1	G	13.9	188	192	189	15.2	8.7
	A566-2[c]	H	9.8	188	190	192	7.9	6.3

[a] Based on the designation of Shoemarker and Specht (1995)
[b] A/A: homozygous PI 97100; B/B: homozygous Coker 237
[c] Putative independent QTL (greater than 50 cM from another marker significantly associated with seed protein and oil, and additive gene action with regards to explaining variation)

pendent QTLs. For two of these loci, the Coker 237 allele was associated with increased protein (A132-4 and A065-1), and for the other two loci the PI 97 100 alleles were also associated with increased protein (Table 4).

The A566-2 locus on LG H accounted for 13.5% of the total phenotypic variation for seed protein. A total of 46.2% of phenotypic variation was explained by these four independent markers. The heritability (assuming a selection unit of two locations, three replications/location) for seed protein in this population was 82.8% (Table 2). Thus, these four markers explained approximately one half the genetic variation for seed protein. This indicated either the existence of other undetected QTLs in this population or else the need to find markers closer to the detected QTLs. The interactions among these four independent markers were not significant ($P \geq 0.01$), suggesting that there was no epistatic effect conditioning seed protein in this population.

Three of four independent markers were detected at both locations, Athens and Blackville, indicating the fairly good agreement of QTLs across locations. The genotype × location variance component for seed protein was 19.2% of the genotypic variance componenet in this population.

Using interval mapping, one QTL in the interval Blt015-1–A132-4 (UNK2) was detected in Athens, and one QTL was exactly on A065-1 (LG K) at Blackville (Table 5). Both the QTL detected in Athens and the QTL detected in Blackville were identified on the combined analysis. When the data were combined, the most likely location of the QTL in the Blt015-1–A132-4 interval was located 10.8 cM from Blt015-1. The most likely position of the QTL on LGK was 11.3 cM from A065-1 toward R051-3 (Table 5). Interval mapping did not detect the QTLs on LG E and H.

The QTL in the integral between R051-3 and N100-1 gave a 3 g/kg increase in seed protein for each allele from

population

	Location	Linkage group[a]	Interval (cM)	Length (cM)	QTL position[b] (cM)	Genetic effects[c]			R^2(%)	LOD[d]
						Additive (a) (g/kg)	Dominant d/a (d) (g/kg)			
protein	Athens	UNK2	Blt015-1 ~ A132-4	11.0	9.2	4		0.4	15.8	3.7
	Blackville	K	A065-1 ~ Q45-1	2.5	0.0	5	−2	−0.4	14.3	3.7
	Combined	K	R051-3 ~ N100-1	17.7	10.2	3	−4	−1.1	14.1	2.7
		UNK2	Blt015-1 ~ A132-4	11.0	10.8	4	2	0.4	13.4	3.4
	Athens	C1	A063-1 ~ Bng143-1	12.4	0.0	−2	1	−0.6	11.4	2.9
		G	cr177-1 ~ L002-1	22.3	16.5	−1	6	−19.2	25.3	4.6
	Blackville	A1	K400-1 ~ B172-1	4.3	3.8	1	−1	−0.6	11.7	2.9
		B1	cr413-1 ~ A469-2	41.2	26.2	3	1	0.3	24.4	2.9
		K	A064-1 ~ R051-3	14.9	8.7	−2	1	−0.5	20.7	4.7
	Combined	G	cr177-1 ~ L002-1	22.3	16.5	1	4	8.2	23.0	4.0

[a] Based on the designation of Shoemaker and Specht (1995)
[b] Most likely QTL position, corresponding to LOD score peak, which represents the distance from left marker of interval
[c] Genetic effects were estimated using Mapmaker/QTL. A sign of negative indicates that the Coker 237 allele decreases the value of the trait
[d] LOD indicates how much more probable the data are to have arisen assuming the presence of a QTL than assuming its absence; LOD threshold = 2.5

Coker 237, and heterozygotes were 4 g/kg lower in seed-protein percentage than expected based on additive gene action (Table 5). At the QTL near A-132-4, each Coker 237 allele was associated with a 4 g/kg increase in seed protein. and the seed protein of heterozygotes was 2 g/kg higher than predicted for additive gene action. The degree of dominance was −1.1 for the QTL in the interval between R051-3 and N100-1, indicating dominant gene action. However, the QTL in the interval between Blt015-1 and A132-4 behaved in a partially dominant manner, with a degree of dominance of 0.4.

Seed oil

Coker 237 averaged 14 g/kg higher seed oil than PI 97 100, and the seed oil content varied by 21 g/kg among progenies (Table 2). On the basis of single-factor ANOVA, five markers were associatd with seed oil content. Three of these five markers on LG C1, G, and H represent putative independent markers (Table 4).

For the A063-1 locus on C1, the PI 97 100 allele increased seed oil content, and the A566-2 locus on LG H was linked to a QTL at which the Coker 237 allele increased seed oil. At the L154-2 locus on LG G, heterozygotes had a higher seed oil percentage than homozygotes, suggesting overdominance or pseudo-overdominance (i.e., repulsion phase linkage of the QTL conditioning oil).

The L154-2 locus explained 17.1% of total phenotypic variation. The amount of variation explained by the three independent markers sums to 40.1%. The heritability (assuming a selection unit of two locations, three replication/location) for seed oil was 74.9% (Table 2). Thus, additional undetected QTLs for seed oil probably exist in this population as well. Two-way ANOVA among these three independent markers did not detect an epistatic effect conditioning seed oil in this population.

Two (A063-1 and A566-2) of the three independent markers were found at both locations. However, the L154-2 locus, which was not detected at Blackville, was identified in the combined analysis over two locations due to its large effect on seed oil at Athens. The other two markers, A235-1 and L002-1, on LG G were detected at both locations. The genotype × location variance component for seed oil was 42% of the genotypice variance component in this population.

Interval mapping revealed two major QTLs on LG C1 and G at Athens (Table 5), which were not detected at Blackville. Instead three major QTLs on LG A1, B1, and K were detected in Blackville. Only one QTL in the interval between cr177-1 and L002-1 on LG G, at exactly the same position as the QTL at Athens, was detected in the combined analysis over two locations. Thus, interval-mapping analysis for seed oil in this population revealed the inconsistency of QTLs across locations, which was in contrast to the results from single-factor ANOVA.

At the QTL in the interval between cr177-1 and L002-1 on LG G, each allele from Coker 237 was associated with increased seed oil, and heterozygotes were 4 g/kg higher in oil content than expected based on additive gene action. As was expected from the higher oil content of heterozygotes for the L154-2 locus in the single-factor ANOVA (Table 4), the degree of dominance at this locus was 8.2, suggesting overdominant gene action. The amount of phenotypic variation explained by this QTL was 23.0%.

Discussion

Relationship between seed protein, seed oil, and maturity

A negative correlation ($r = -0.61$) was found for protein and oil percentage in the PI 97 100 × Coker 237 population, which is in agreement with earlier studies (Johnson and Bernard 1962; Johnson et al. 1955; Kwon and Torrie 1964; Smith and Weber 1968). In previous mapping studies, the association between these two traits was explained by QTLs conditioning both traits (Diers et al. 1992; Mansur et al. 1993). For seed protein and oil, Diers et al. (1992) reported common RFLP loci clustered on LG E and I (formerly described as LG A and K, respectively). Of the 14 loci for seed protein and the 15 loci for seed oil that were significant at the 0.05 probability level (data not shown) using single-factor ANOVA in the PI 97 100 × Coker 237 population, we found various marker loci associated with both protein and oil. They were A454-1 on LG E, A235-1, L002-1, and L154-2 on LG G, A566-2 on LG H, R051-3 on LG K, and A132-4 on LG UNK2. At each of these RFLP loci the allele associated with increased seed protein was associated with decreased seed oil (data not shown). The QTLs near these markers could provide the genetic basis for the negative correlation between seed protein and oil.

However, a correlation was not found between seed protein and oil in the Young × PI 416 937 population ($r = -0.05$). Of the 29 loci for seed protein and the 15 loci for seed oil at the level of $P = 0.05$, only A069-3 and cr167-1 on LG E and K258-1 on LG R were associated with both seed protein and oil. This paucity of common markers for seed protein and oil in the Young × PI 416 937 would explain the lack of association between seed protein and oil in this population, but does not preclude the possibility that existing markers associated with both traits were not uncovered in this study.

Protein and oil content of soybean have been previously associated with maturity. Depending on the population studied, there were negative or positive associations between seed protein and maturity (Weiss et al. 1952; Kwon and Torrie 1964; Simpson and Wilcox 1983). A negative correlation between maturity and seed protein content ($r = -0.45^{***}$) was present in the

Young × PI 416 937 population, whereas no association was present in the PI 97 100 × Coker 237 population. This might be due to the 35-day range of maturity in the Young × PI 416 937 population compared to the relatively small 10-day range in the PI 97 100 × Coker 237 population. In the Young × PI 416 937 population, one marker (gac 197-1 on LG C1) was associated with both seed protein and maturity. At this locus the PI 416 937 allele for increased protein was associated with earlier maturity. Although there was no association between maturity and seed protein in the PI 97 100 × Coker 237 population, the R051-3 marker on LG K was associated with both seed protein and maturity. At this locus the Coker 237 allele conditioned earlier maturity and increased seed protein content. The other marker for maturity on LG K (B032-2) was not associated with seed protein in this population. These loci, gac197-1 and R051-3, may provide the basis that maturity affects the seed protein content, suggesting the utility of the other markers identified in this study in a marker-assisted breeding program for high seed protein content.

Contrary to the previous reports of the consistent association between high oil content and early maturity (Weber and Moorthy 1952; Johnson et al. 1955; Kwon and Torrie 1964; Simpsion and Wilcox 1983), these traits were not closely associated in our two populations. However, one locus on LG L (A023-1) associated with seed oil in the Young × PI 416 937 population was 3.2 cM away from the O109-1 locus associated with maturity. At this locus the PI 416 937 allele for increased seed oil was associated with earlier maturity.

Consitency of QTLs across locations

Using single-factor ANOVA, all four putative independent markers for seed oil in the Young × PI 416 937 population and three of four for seed protein in the PI 97 100 × Coker 237 population were detected in each location. Also, five of seven independent markers for seed protein in the Young × PI 416 937 population were detected at all three locations, and two of three for seed oil in the PI 97 100 × Coker 237 were detected in the two locations of this experiment. This indicated a fairly good agreement of QTLs across locations. Stuber et al. (1992) reported little evidence for genotype-by-location interaction for most QTLs in maize lines grown in six diverse environments.

A consistency of QTLs associated with plant height, lodging, and maturity was also reported by Lee et al. (1996b) for soybean. They surmised that the consistency of QTLs across environments resulted from the control of the traits by a few loci with large effects. However, many other studies (Paterson et al. 1991; Bubeck et al. 1993; Lee et al. 1996a) revealed inconsistency of some QTLs over environments. These differences may be due to the heritability of the traits evaluated, the number of QTLs segregating, the level of probability for significance tests, sampling error, and the specific crop species. In our study, heritabilities were fairly high (75 to 89%) for seed protein and oil (Table 2), and our probability level ($P \leqslant 0.05$) was chosen to enhance the ability to detect QTL consistency.

In contrast to single-factor ANOVA, interval mapping in the PI 97 100 × Coker 237 population revealed an inconsistency of QTLs across locations. The QTLs at Athens for seed protein and oil were totally different from those at Blackville (Table 5). These differences may also be due to the low level of saturation of the genetic map.

Population specificity

Unique QTLs for seed protein and seed oil were determined on the basis of the position on different LGs as well as the genetic map distance on the same LG. Only one QTL on LG E was consistently associated with seed protein in both populations. However, none of QTLs associated with seed oil were common in both populations.

Comparisons of QTLs for seed protein and oil identified in this study were made with those reported by Diers et al. (1992) and Mansur et al. (1993) to identify unique QTLs. In F_2-derived lines from the cross between *G. max* and *G. soja*, Diers et al. (1992) identified eight RFLP markers on LG E, H, and I (formerly described as LG A, C, and K, respectively) associated with seed protein. For seed oil, nine markers were located on LG E and I of Shoemaker and Specht (1995). Mansur et al. (1993) reported that an unlinked RFLP locus, L48, was associated with seed protein in an F_2-derived soybean population from Minsoy × Noir 1. For seed oil,two QTLs near T153a (LG A2, Shoemaker and Specht 1995) and A315 (possibly LG K) were identified. On the basis of LG K of the USDA map (Shoemaker and Specht 1995), the locus A065-1 associated with seed protein in the PI 97 100 × Coker 237 population (Table 4) is 23 cM away from the A315-1 locus which was associated with seed oil in the study by Mansur et al. (1993).

Several markers on LG E were associated with seed protein and oil in our study. These include A517-1 and cr167-1 for seed protein, and A069-3 for seed oil in the Young × PI 416 937 population (Table 3), and A454-1 for seed protein in PI 97 100 × Coker 237 (Table 4). Also, Diers et al. (1992) found QTLs for protein and oil on LG E. To determine if the QTLs mapped on LG E among these three populations were similar, the relative QTL positions were drawn in Fig. 1 on the basis of LG E of the USDA map (Shoemaker and Specht 1995). The A517-1 locus for seed protein in the Young × PI 416 937 population and A454-1 in the PI 97 100 × Coker 237 population were 12.4 and 18.0 cM away from the *pb* locus, respectively. This locus was associated with seed protein in the study of Diers et al. (1992). These two markers are in the same position as the QTL for seed oil in the genomic region between K229-1 and A203-1

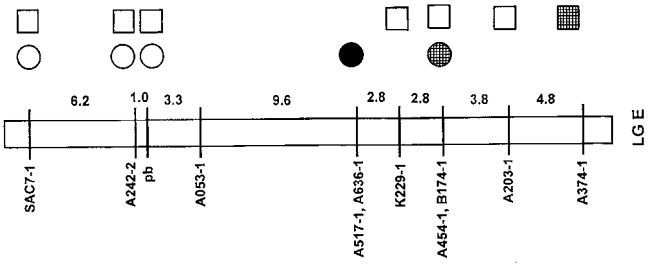

Fig. 1 Comparison of QTLs for seed protein and oil on LG E derived from two populations, Young × PI 416937 and PI 97100 × Coker 237, with those identified by Diers et al. (1992). The map was adapted from the part of LG E in USDA map (Shoemaker and Specht 1995). The *open*, *filled*, and *hatched circles* represent the position of QTLs associated with seed protein for *G. Max* × *G. soja* (Diers et al. 1992), Young × PI 614937, and PI 97100 × Coker 237 populations, respectively, and *squares* represent the position of QTLs associated with seed oil

detected by Diers et al. (1992). Also, the A069-3 locus was associated with seed oil in Young × PI 416937. A069-3 is near A374-1 on the USDA map and is within 5 cM from A203-1 which is associated with seed oil (Diers et al. 1992).

Except for those QTLs associated with seed protein and oil on LG E, there were new genomic locations for putative QTLs identified on LG C1, H, J, K, N, P, and at least two unknown linkage groups for protein, and LG C1, G, H, J, L, and R for oil in these populations. Comparison of our study with those of others emphasized the population specificity of important QTLs for seed protein and oil, and provided evidence for additional QTLs for seed protein and oil in soybean.

Acknowledgements The authors wish to thank Barbara Stewart for technical assistance and R. C. Shoemaker, K. G. Lark, R. T. Nagao, G. D. Kochert, N. D. Young and J. M. Tohme for providing RFLP probes. Suk-Ha Lee thanks the International Atomic Energy Agency for financial assistance in the form of an IAEA fellowship. This research was funded by state and Hatch funds allocated to the Georgia Agricultural Experimental Stations and by a grant from the United Soybean Board.

References

Brim CA, Burton JW (1979) Recurrent selection in soybeans. II. Selection for increased percent protein in seeds. Crop Sci 19:494–498

Burr B, Burr FA, Thompson KH, Albertson MC, Stuber CW (1988) Gene mapping with recombinant inbreds in maize. Genetics 118:519–526

Burton JW (1985) Breeding soybeans for improved protein quantitiy and quality. In: Shibles R (ed) Proc 3rd World Soybean Res Conf, Westview Press, Boulder, Colorado, pp 361–367

Burton JW, Brim CA (1981) Recurrent selection in soybeans. III. Selection for increased percent oil in seeds. Crop Sci 21:31–34

Burton JW, Brim CA, Young MF (1987) Registration of Young soybean. Crop Sci 27:1093

Bubeck DM, Goodman MM, Beavis WD, Grant D (1993) Quantitative trait loci controlling resistance to gray leaf spot in maize. Crop Sci 33:838–847

Byth DE, Weber CR, Caldwell BE (1969) Correlated truncation selection for yield in soybeans. Crop Sci 9:699–702

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

Dudley JW (1993) Molecular markers in plant improvement: manipulation of genes affecting quantitative traits. Crop Sci 33:660–668

Dudley JW (1994) Comparison of genetic distance estimators using molecular marker data. In: Analysis of molecular marker data. Joint Plant Breeding Symposia Services August 1994 Corvallis Oregon. Am Soc for Hort Sci and Crop Sci Soc of Amer, pp 1–7

Fehr WR, Weber CR (1968) Mass selection by seed size and specific gravity in soybean populations. Crop Sci 8:551–554

Haldane JBS, Waddington CH (1931) Inbreeding and linkage. Genetics 16:357–374

Holloway JL and Knapp SJ (1993) Gmendel 3.0 users guide. Department of Crop and Soil Sciences, Oregon State University

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

Johnson HW, Robinson HF, Comstock RE (1955) Genotypic and phenotypic correlations in soybeans and their implications in selection. Agron J 47:477–483

Keim P, Diers BW, Shoemaker RC (1990a) Genetic analysis of soybean hard seededness with molecular markers. Theor Appl Genet 79:465–469

Keim P, Diers BW, Olson TC, Shoemaker RC (1990b) RFLP mapping in soybean: association between marker loci and variation in quantitative traits. Genetics 126:735–742

Kwon SH, Torrie JH (1964) Heritability of and interrelationships among traits of two soybean populations. Crop Sci 4:196–198

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

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

Lark KG, Weisemann JM, Matthews BF, Palmer R, Chase K, Macalma T (1993) A genetic map of soybean (*Glycine max* L.) using an intraspecific cross of two cultivars: 'Minsoy' and 'Noir 1'. Theor Appl Genet 86:901–906

Lee SH, Bailey MA, Mian MAR, Carter Jr TE, Ashley DA, Hussey RS, Parrott WA, Boerma HR (1996a) Molecular markers associated with soybean plant height, lodging, and maturity across locations. Crop Sci 36:728–735

Lee SH, Bailey MA, Mian MAR, Shipe ER, Ashley DA, Parrott WA, Hussey RS, Boerma HR (1996b) Identification of quantitiative trait loci for plant height, lodging, and maturity in a soybean population segregating for growth habit. Theor Appl Genet 92:516–523

Lincoln S, Daly M, Lander E (1992) Mapping genes controlling quantitative traits with MAPMAKER/QTL. Whitehead Institute Tech Rep, 2nd edn.

Mansur LM, Lark KG, Kross H, Oliveria A (1993) Interval mapping of quantitative trait loci for reproductive, morphological, and seed traits of soybean (*Glycine max* L.). Theor Appl Genet 86:907–913

Mian MAR, Bailey MA, Ashley DA, Wells R, Carter Jr TE, Parrott WA, Boerma HR (1996) Molecular markers associated with water use efficiency and leaf ash in soybean. Crop Sci 36: Sept–Oct issue

Paterson AH, Damon S, Hewitt JD, Zamir D, Rabinowitch HD, Lincoln SE, Lander ES, Tanksley SD (1991) Mendelian factors underlying quantitative traits in tomato: comparison across species, generations, and environments. Genetics 127:181–197

SAS (1988) SAS/STAT User's Guide, version 6.03. SAS Institute, Cary, North Carolina

Shoemaker RC, Specht JE (1995) Integration of the soybean molecular and classical genetic linkage groups. Crop Sci 35:436–446

Simpson AM, Jr, Wilcox JR (1983) Genetic and phenotypic associations of agronomic characteristics in four high-protein soybean populations. Crop Sci 23:1077–1081

Sloane RJ, Patterson RP, Carter Jr TE (1990) Field drought tolerance of a soybean plant introduction. Crop Sci 30:118–123

Smith RR, Weber CR (1968) Mass selection by specific gravity for protein and oil soybean populations. Crop Sci 8:373–377

Stuber CW, Lincoln SE, Wolff DW, Helentjaris T, Lander ES (1992) Identification of genetic factors contributing to heterosis in a hybrid from two elite inbred lines using molecular markers. Genetics 132:823–839

Tanksley SD, Young ND, Paterson AH, Bonierbale MW (1989) RFLP mapping in plant breeding: new tools for an old science. Bio/Technology 7:257–264

Weber CR, Moorthy (1952) Heritable and nonheritable relationships and variability of oil content and agronomic characteristics in the F_2 generation of soybean crosses. Agron J 44:202–209

Weiss MG, Weber CR, Williams LF, Probst AH (1952) Correlation of agronomic characters and temperature with seed compositional characters in soybeans, as influenced by variety and time of planting. Agron J 44:289–297

Wilcox JR (1985) Breeding soybeans for improved oil quantity and quality. In: Shibles R (ed) Proc 3rd World Soybean Res Conf, Westview Press, Boulder, Colorado, pp 380–386

Xiao J, Li J, Yuan L, Tanksley SD (1995) Dominance is the major genetic basis of heterosis in rice as revealed by QTL analysis using molecular markers. Genetics 140:745–754