Theor Appl Genet (1992) 83:608–612

THEORETICAL AND APPLIED GENETICS

RFLP analysis of soybean seed protein and oil content

B. W. Diers [1], P. Keim [2], W. R. Fehr [1], and R. C. Shoemaker [3, 4, *]

[1] Department of Agronomy, Iowa State University, Ames, IA 50011, USA
[2] Department of Biology, Northern Arizona University, Flagstaff, AZ 86011, USA
[3] Departments of Agronomy and Genetics, Iowa State University, Ames, IA 50011, USA
[4] USDA-Agricultural Research Service, Field Crops Research Unit, Iowa State University, Ames, IA 50011, USA

Received May 25, 1990; Accepted July 9, 1991
Communicated by A. L. Kahler

Summary. The objectives of this study were to present an expanded soybean RFLP map and to identify quantitative trait loci (QTL) in soybean [*Glycine max* (L.) Merr.] for seed protein and oil content. The study population was formed from a cross between a *G. max* experimental line (A81-356022) and a *G. soja* Sieb. and Zucc. plant introduction (PI 468916). A total of 252 markers was mapped in the population, forming 31 linkage groups. Protein and oil content were measured on seed harvested from a replicated trial of 60 F_2-derived lines in the F_3 generation ($F_{2:3}$ lines). Each $F_{2:3}$ line was genotyped with 243 RFLP, five isozyme, one storage protein, and three morphological markers. Significant ($P<0.01$) associations were found between the segregation of markers and seed protein and oil content. Segregation of individual markers explained up to 43% of the total variation for specific traits. All *G. max* alleles at significant loci for oil content were associated with greater oil content than *G. soja* alleles. All *G. soja* alleles at significant loci for protein content were associated with greater protein content than *G. max* alleles.

Key words: Restriction fragment length polymorphism (RFLP) – *Glycine max* – Quantitative trait loci (QTL) – Protein – Oil

Introduction

Soybean is grown primarily for the protein and oil processed from its seed (Smith and Huyser 1987). The increasing interest in soybean genotypes that fit into specific markets and competition from other oil seed crops will continue to make increased seed protein and oil major breeding objectives in the future. Both protein and oil content are quantitatively inherited in soybean (Burton 1985; Wilcox 1985). Breeders have been successful in manipulating these traits, but their underlying genetic controls have not been elucidated.

Genetic markers have allowed researchers to systematically map and characterize genes that are important in conferring quantitative traits. These genes have been mapped to what has become known as quantitative trait loci (QTL). The use of restriction fragment length polymorphism (RFLP) markers has increased the efficiency of mapping QTLs, because of the greater number of markers that can be scored in a single population relative to other markers used such as isozyme or morphological markers. Genetic mapping of QTL has been documented in maize and tomato (Edwards et al. 1987; Osborn et al. 1987; Paterson et al. 1988). QTL mapping has led to an increased understanding of genes involved in the inheritance of quantitative traits, and may improve genetic gains in breeding programs through marker-assisted selection.

Molecular markers have been used to identify QTL in *G. max* × *G. soja*-derived populations. Graef et al. (1989) and Suarez (1989) studied the association between isozyme markers and quantitative traits in two *G. soja* × *G. max* backcross populations. Both found significant associations between vegetative traits and isozyme loci. Suarez found several associations between specific isozyme markers and protein, oil, and fatty acid content; however, most of the associations were population specific. Their studies were limited by the low number of polymorphic isozyme markers available in their populations: six in one and eight in the other.

Keim et al. (1990a) studied hard seededness in a *G. max* × *G. soja* single-cross population with 70 RFLP

* To whom correspondence should be addressed.

markers. QTL were found that explained 71% of the total variation for hard seededness in the population. Keim et al. (1990 b) continued mapping in the same population and found significant QTL for maturity and morphological traits by using 150 RFLP markers. They found markers that explained more than 20% of the total variation for several traits.

The purpose of this research was to use RFLP technology to map quantitative trait loci for seed protein and oil content in the same population used by Keim et al. (1990 a, b).

Materials and methods

The study was conducted with a population produced from a cross between the *G. max* experimental line A81-356022 and the *G. soja* accession PI 468916. This population was used to develop the public RFLP map, and F_2 data from all of the markers were used to map QTL.

Sixty F_2 plants from the mapping population were grown at the Iowa State University Agronomy and Agricultural Engineering Research Center near Ames/IA, during the summer of 1987. Leaf samples were taken from each plant for DNA extraction and RFLP analysis. The plants were allowed to naturally self-pollinate and, at maturity, each plant was harvested and threshed separately to form F_2-derived ($F_{2:3}$) lines.

Linkage mapping

A total of 252 loci was scored in the population to construct the linkage map. Two hundred and twenty-seven of the loci were scored using low-copy clones from a *PstI* genomic library of soybean (Keim and Shoemaker 1988). Also included on the map were 16 loci scored using recombinant DNA clones obtained from other labs (see Fig. 1 legend), five isozyme markers, three morphological markers, and one storage protein marker. The DNA extraction, Southern blotting, and hybridization procedures have been described elsewhere (Keim et al. 1989). The linkage map from the F_2 segregation data was constructed using the program Mapmaker (Lander et al. 1987). A minimum lod score of 3.0 was used, with the exception of the linkage of markers pA-203 and pT-153b, where a lod of 2.8 was used.

QTL mapping

The $F_{2:3}$ lines and parents were evaluated during the summer of 1988 in a randomized complete-block design experiment with two replications at each of three locations near Ames/IA. The locations were the Agronomy and Agricultural Engineering Research Center, the Burkey Farm, and the Bruner Farm. Plots were single rows 1.5 m long, with 1-m row spacing and a seeding rate of 33 seeds m^{-1}. Plots at the Burkey Farm were planted 1 May, at the Agronomy and Agricultural Engineering Research Center on 15 May, and at the Bruner Farm on 29 May. Each plot was harvested and threshed separately at maturity. Seed protein and oil content were measured from a 5- to 7-g ground sample from each plot at the USDA Northern Regional Research Center at Peoria/IL, by using a Pacific-Scientific NIR grain analyzer.

Seed trait data were analyzed by standard analysis of variance procedures for a randomized complete-block design model. Variance component estimates and broad-sense heritabilities were calculated according to Fehr (1987). Each marker-seed trait combination was analyzed to determine if segregation of individual markers explained significant seed trait variation. The lines were divided into three classes for codominant markers (homozygous for *G. max* alleles, homozygous for *G. soja* alleles, and heterozygous) or two classes for dominant markers (heterozygous class and homozygous dominant class contrasted with the homozygous recessive class). A single-factor analysis of variance was used to determine if significant differences were present among marker classes. Significance was determined by *F*-tests. The amount of variation explained by a marker was determined by using the R^2 value, which is the proportion of the total variance among the 60 line means explained by the segregation of a marker.

Results

The RFLP data were used to construct a RFLP linkage map of soybean (Fig. 1). The map contains 252 markers, 31 linkage groups, and 2,147 centiMorgans (cM). The average distance between two adjacent marker loci is 8.5 cM. Several linkage groups still must be joined because the soybean haploid genome contains 20 chromosomes (Palmer and Kilen 1987). Twenty-five of the clones gave hybridization patterns that allowed two loci to be scored. Three clones gave patterns that allowed three loci to be scored. Refer to Keim et al. (1990 b) for a more detailed discussion on the linkage mapping results.

The *G. max* and *G. soja* parents were significantly different ($P<0.001$) for seed protein and oil content based on analysis of seed from the field trial. The *G. max* and *G. soja* parents contained 420 and 471 g $(kg\ seed)^{-1}$ protein and 198 and 101 g $(kg\ seed)^{-1}$ oil, respectively. Significant genetic variation was present among the $F_{2:3}$ lines for both traits. The broad-sense heritability was 0.74 for protein and 0.92 for oil content. Each marker-seed trait combination was then tested to determine if significant associations existed between the segregation of markers and variation for the traits. Markers that were associated with significant variation for protein and

Fig. 1. Soybean RFLP map. The tentative names of the linkage groups are listed at the top of each group. The markers labelled pA and pK are RFLP markers developed by the USDA-ARS at Iowa State. Markers labelled M13, pR, pT pG, and PC were developed by Dr. K. G. Lark at the University of Utah. Markers labelled pSAC are actin gene probes kindly provided by Dr. R. Meagher from the University of Georgia. Included also on the map are five isozyme markers (Rennie et al. 1989), diaphorase (*Dia*), isocitrate dehydrogenase (*Idh*) (Palmer and Kilen 1987) and acid phosphatase (ACP), malic enzyme (ME), and malate dehydrogenase (MDH), the *Gy4* storage protein locus (Palmer and Kilen 1987) and the morphological markers *i* (seed coat color), *pb* (pubescent tip) (Palmer and Kilen 1987), and SCD (seed coat luster)

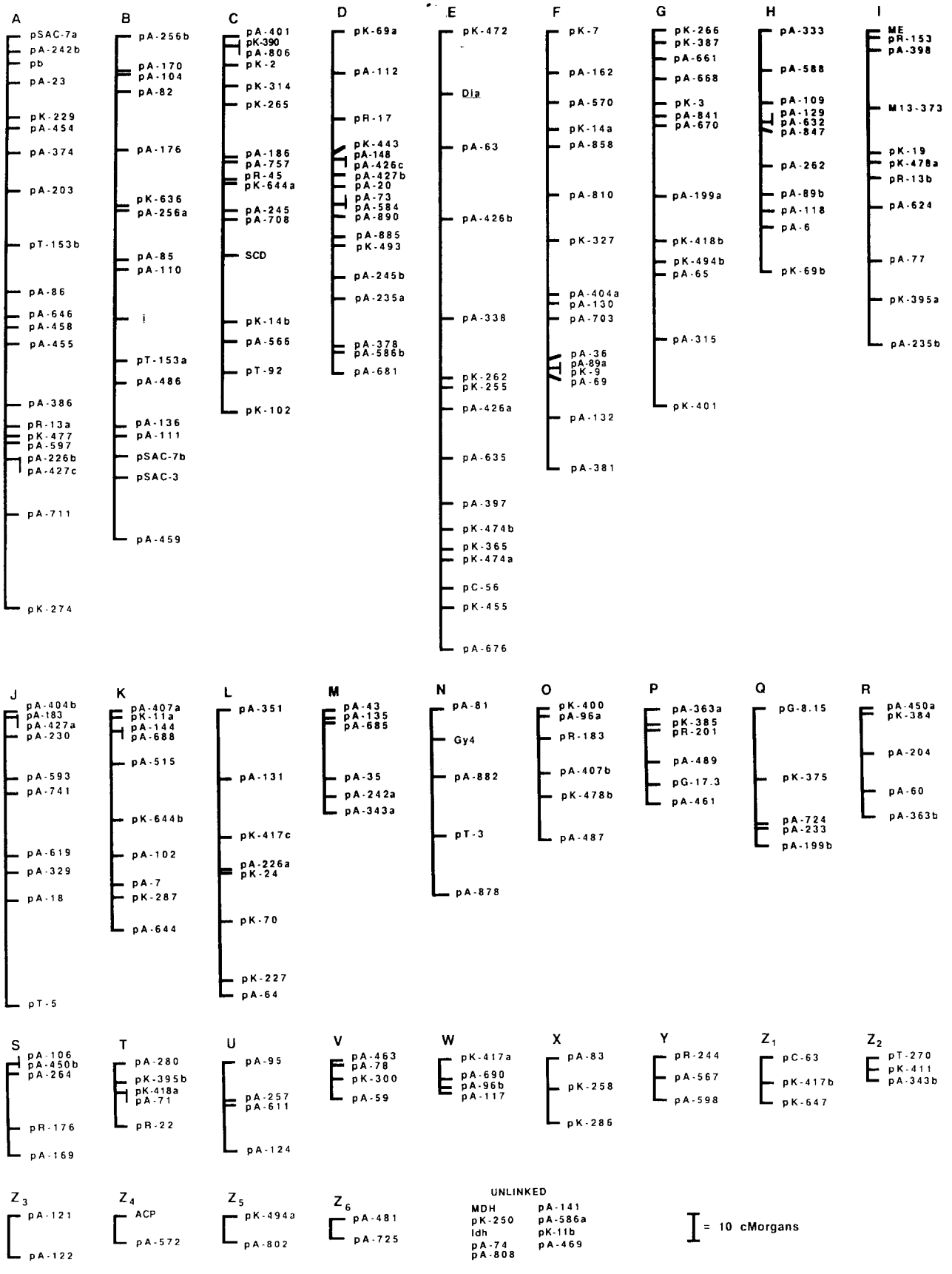
A
pSAC-7a
pA-242b
pb
pA-23
pK-229
pA-454
pA-374
pA-203
pT-153b
pA-86
pA-646
pA-458
pA-455
pA-386
pR-13a
pK-477
pA-597
pA-226b
pA-427c
pA-711
pK-274
B
pA-256b
pA-170
pA-104
pA-82
pA-176
pK-636
pA-256a
pA-85
pA-110
pT-153a
pA-486
pA-136
pA-111
pSAC-7b
pSAC-3
pA-459
C
pA-401
pK-390
pA-806
pK-2
pK-314
pK-265
pA-186
pA-757
pR-45
pK-644a
pA-245
pA-708
SCD
pK-14b
pA-566
pT-92
pK-102
D
pK-69a
pA-112
pR-17
pK-443
pA-148
pA-426c
pA-427b
pA-20
pA-73
pA-584
pA-890
pA-885
pK-493
pA-245b
pA-235a
pA-378
pA-586b
pA-681
E
pK-472
Dia
pA-63
pA-426b
pA-338
pK-262
pK-255
pA-426a
pA-635
pA-397
pK-474b
pK-365
pK-474a
pC-56
pK-455
pA-676
F
pK-7
pA-162
pA-570
pK-14a
pA-858
pA-810
pK-327
pA-404a
pA-130
pA-703
pA-36
pA-89a
pK-9
pA-69
pA-132
pA-381
G
pK-266
pK-387
pA-661
pA-668
pK-3
pA-841
pA-670
pA-199a
pK-418b
pK-494b
pA-65
pA-315
pK-401
H
pA-333
pA-588
pA-109
pA-129
pA-632
pA-847
pA-262
pA-89b
pA-118
pA-6
pK-69b
I
ME
pR-153
pA-398
M13-373
pK-19
pK-478a
pR-13b
pA-624
pA-77
pK-395a
pA-235b
J
pA-404b
pA-183
pA-427a
pA-230
pA-593
pA-741
pA-619
pA-329
pA-18
pT-5
K
pA-407a
pK-11a
pA-144
pA-688
pA-515
pK-644b
pA-102
pA-7
pK-287
pA-644
L
pA-351
pA-131
pK-417c
pA-226a
pK-24
pK-70
pK-227
pA-64
M
pA-43
pA-135
pA-685
pA-35
pA-242a
pA-343a
N
pA-81
Gy4
pA-882
pT-3
pA-878
O
pK-400
pA-96a
pR-183
pA-407b
pK-478b
pA-487
P
pA-363a
pK-385
pR-201
pA-489
pG-17.3
pA-461
Q
pG-8.15
pK-375
pA-724
pA-233
pA-199b
R
pA-450a
pK-384
pA-204
pA-60
pA-363b
S
pA-106
pA-450b
pA-264
pR-176
pA-169
T
pA-280
pK-395b
pK-418a
pA-71
pR-22
U
pA-95
pA-257
pA-611
pA-124
V
pA-463
pA-78
pK-300
pA-59
W
pK-417a
pA-690
pA-96b
pA-117
X
pA-83
pK-258
pK-286
Y
pR-244
pA-567
pA-598
Z1
pC-63
pK-417b
pK-647
Z2
pT-270
pK-411
pA-343b
Z3
pA-121
pA-122
Z4
ACP
pA-572
Z5
pK-494a
pA-802
Z6
pA-481
pA-725
UNLINKED
MDH
pK-250
Idh
pA-74
pA-808
pA-141
pA-586a
pK-11b
pA-469
= 10 cMorgans

Table 1. Markers significantly ($P < 0.01$) associated with variation for protein and oil content

Marker	R^2	$P > F$	Means of genotypic classes[a]			Linkage group
			MM	SM	SS	
			g (kg seed)$^{-1}$			
Protein						
pK-11a	0.42	0.0001	450	465	474	K
pA-407a	0.39	0.0001	451	464	473	K
pA-144	0.24	0.0007	454	464	472	K
pA-688	0.25	0.001	454	465	472	K
pSAC-7a[b]	0.24	0.003	455	464	472	A
pA-242b	0.19	0.004	456	465	468	A
pA-23	0.16	0.01	457	466	468	A
pA-245a[c]	0.12	0.01	455		465	C
Total oil						
pSAC-7a	0.43	0.0001	153	145	136	A
pA-242b	0.39	0.0001	154	143	140	A
pA-23	0.32	0.0001	153	144	139	A
pb[d]	0.27	0.0001	152	144	139	A
pK-11A	0.27	0.0002	155	144	140	K
pA-407a	0.28	0.0005	154	144	140	K
pA-454	0.23	0.0008	152	143	142	A
pK-229	0.22	0.001	151	144	140	A
pA-203	0.18	0.006	152	145	142	A

[a] MM designates homozygous *G. max* class, SM heterozygous class, and SS homozygous *G. soja*
[b] Actin gene probe provided by Dr. R. Meagher (University of Georgia)
[c] *G. soja* alleles dominant to *G. max* alleles
[d] Morphological marker blunt-sharp pubescence tip (Palmer and Kilen 1987)

oil at $P<0.01$ are given in Table 1. Some markers that were associated with significant variation for a given trait mapped to the same linkage group, suggesting that these markers were associated with variation controlled by the same gene or group of genes (Table 1).

All *G. soja* alleles at loci significant for protein were associated with greater protein content than *G. max* alleles. All *G. max* alleles at significant loci for oil were associated with greater oil content than *G. soja* alleles. These results were expected because the *G. max* parent had greater oil and lesser protein content than the *G. soja* parent. The significant marker loci for protein and oil content were mostly clustered on linkage groups 'A' and 'K,' which suggested that important genes for these traits were located within these linkage groups. Alleles significantly associated with increased levels of one trait, protein or oil, generally were associated with lower levels of the other trait (Table 1). This is consistent with the negative correlation generally found between these traits in soybean (Burton 1985).

Discussion

A soybean linkage map consisting of 252 markers has been presented in this paper. Markers used in the mapping were associated with regions of the genome that explained significant variation for protein and oil content. These regions likely contain important genes that control expression of the seed traits. The large amount of variation associated with individual markers in this population suggested that the inheritance of these traits was controlled partly by genes with large effects. However, of the relatively small population that was used in this study, it is likely that only those loci with alleles having a relatively large effect on the quantitative traits were identified. There are likely many more loci that affect protein and oil content segregating in the population, but their effects were too minor to be identified in these analyses.

There is a probability that some markers were erroneously declared significant (Type I errors). With the probability cutoff of 0.01 used in this experiment, the chance of at least one Type I error occurring among the 252 analyses of variance calculated for each trait is 0.92 (Lander and Botstein 1989). The probability of a Type I error decreases to 0.22 for a probability level of 0.001 and to 0.025 for a probability level of 0.0001. Although the chance of Type I errors is high with use of a probability of 0.01, the loci associated with the largest trait variation were significant at probability values as low as 0.0001 (Table 1).

It should be noted that the *Gy4* gene (Palmer and Kilen 1987), which encodes the G4 subunit of the glycinin storage protein, was mapped in this population (Fig. 1). *Gy4* was mapped using the segregation of a RFLP revealed using a clone of *Gy4* as a probe (Dickinson et al. 1987), and the segregation of a G4 polypeptide, A4. The *G. soja* parent had a lower mobility for the polypeptide A4 than previously observed in soybean germplasm. This allowed the scoring of the segregation of A4 using polyacrylamide gel electrophoresis. The RFLP and the storage protein variant co-segregated among the F_2s. These data were used to test for an association between the storage protein gene and total protein content in the seed. No significant association was found, which suggests that this structural storage protein gene or a closely linked gene does not have a major effect on protein content in this population.

Studies are being initiated that will test for RFLP associations with protein and oil content in a number of other populations. Information on the association between markers and protein and oil content should help breeders in constructing allelic combinations for further study and for the development of superior genotypes.

Acknowledgements. The authors acknowledge the work of Terry Olson, who collected most of the RFLP data and developed the

map presented in this manuscript. This is Journal Paper No. J-14016 of the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa. Project No. 2763. This work was supported, in part, by Iowa State Biotechnology Grant No. 480-46-09.

References

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

Dickinson CD, Floener LA, Lilley GG, Nielsen NC (1987) Self-assembly of proglycinin and hybrid proglycinin synthesized in vitro from cDNA. Proc Natl Acad Sci USA 84:5525–5529

Edwards MD, Stuber CW, Wendel JF (1987) Molecular-marker-facilitated investigations of quantitative trait loci in maize. I. Numbers, genomic distribution and types of gene action. Genetics 116:113–125

Fehr WR (1987) Principles of cultivar development. Macmillan, New York

Graef GL, Fehr WR, Cianzio SR (1989) Relation of isozyme genotypes to quantitative characters in soybean. Crop Sci 29:683–688

Keim P, Shoemaker RC (1988) Construction of a random recombinant DNA library that is primarily single copy sequences. Soybean Genet Newsl 15:147–148

Keim P, Shoemaker RC, Palmer RG (1989) Restriction fragment length polymorphism diversity in soybean. Theor Appl Genet 77:786–797

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 T, Shoemaker RC (1990b) RFLP mapping in soybean: association between marker loci and variation in quantitative traits. Genetics 126:735–742

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

Osborn TC, Alexander DC, Fobes JF (1987) Identification of restriction fragment length polymorphisms linked to genes controlling soluble solids content in tomato fruit. Theor Appl Genet 73:350–356

Palmer RG, Kilen TC (1987) Qualitative genetics and cytogenetics. In: Wilcox JR (ed) Soybean: improvement, production and uses, 2nd edn. Agronomy 16:23–48

Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721–726

Rennie BD, Thorpe ML, Beversdorf WD (1989) A procedural manual for the detection and identification of soybean [*Glycine max* (L.) Merr.] isozymes using the starch gel electrophoretic system. Dept Crop Sci Tech Bull. University of Guelph, Ontario, Canada

Smith KJ, Huyser W (1987) World distribution and significance of soybean. In: Wilcox JR (ed) Soybeans: improvement, production, and uses, 2nd edn. Agronomy 16:23–48

Suarez JC (1989) Association between isozyme genotypes with quantitative traits in soybean. PhD thesis, Iowa State University, Ames/IA (Diss Abstr 89-20190)

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