Theor Appl Genet (1990) 79:465–469

Genetic analysis of soybean hard seededness with molecular markers*

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

[1] Departments of Agronomy and Genetics, Iowa State University, Ames, IA 50011, USA
[2] United States Department of Agriculture-Agricultural Research Service, Agronomy Hall, Ames, IA 50011, USA

Received June 23, 1989; Accepted November 2, 1989
Communicated by K. Tsunewaki

Summary. Hard seededness in soybean [*Glycine max* (L.) Merr.] is a quantitative trait that affects the germination rate, viability, and quality of stored seeds. We have used 72 restriction fragment length polymorphisms (RFLPs) to identify genomic regions containing quantitative trait loci (QTL) affecting hard seededness in a segregating population from a *G. max* by a *Glycine soja* (Sieb. & Zucc.) cross. Five independent RFLP markers were found to be associated with variation in the hard-seeded trait. These markers and the epistatic interactions between them explain 71% of the variation for hard seededness. A genomic region associated with the *i* locus accounted for 32% of the variation in this segregating population. This study illustrates one approach to physiological genetic studies in plants.

Key words: Soybean – RFLP – Quantitative trait loci – Germination – Hard seededness

Introduction

Imbibition is one of the first steps in breaking dormancy and initiating germination of a soybean seed. Dormancy and viability can be maintained for long periods in hard-seeded soybean accessions because their seed coats are impermeable to water (Rolston 1978). Some legume seeds have remained viable for more than 100 years (Rolston 1978). This is important ecologically in wild populations and economically in cultivated legumes. Typically, wild soybean accessions are hard seeded, whereas cultivated varieties are not. In the southern United States and in the tropics, hard seededness would be beneficial by contributing to the quality of stored soybeans and to their viability (Potts et al. 1978). Some breeding programs have recognized the importance of hard seededness and have introgressed this trait into adapted cultivars (Kilen and Hartwig 1978).

The physiological basis for seed-coat impermeability is not fully understood (Rolston 1978, Ballard 1973), but morphological (Egley and Paul 1981; Yaklich et al. 1984), enzymatic (Egley et al. 1983), and phenolic differences (Marbach and Mayer 1974, 1975) have been implicated. Even though as few as three genes may control variation in soybean hard seededness in some populations (Kilen and Hartwig 1978), the relationship between the physiology and the genetics of hard seededness has not yet been explored.

The use of genetics to understand physiological phenomena has been invaluable in bacteria and fungi (Gots and Benson 1974; Henry et al. 1984), and this success has fostered many attempts to develop comparable systems in plants (Terzaghi et al. 1985). The problems associated with these types of studies in plants have been enormous, and plant genetics has played a relatively minor role in physiological studies when compared with other systems. *Arabidopsis* seems the best candidate for integrating genetic and physiological studies, and notable success has been achieved (Bowman et al. 1989). However, this success is dependent upon the unique biology of *Arabidopsis* and may not be generally applicable to studies in other plants. Novel genetic strategies will be required for wide application of physiological-genetic studies in soybean.

Restriction fragment length polymorphisms (RFLPs) as genetic markers have facilitated the genetic study of traits conditioned by multiple genes (Burr et al. 1989).

* Journal Paper No. J-13557 of the Iowa Agriculture and Home Economics Experiment Station, Ames, IA, Project No. 2763
** Current address: Department of Biology, NAU Box 5640, Northern Arizona University, Flagstaff, AZ 86011, USA

Marker loci genetically linked to quantitative trait loci (QTLs) can be used to first identify genomic regions that contain important genes, and then to facilitate the genetic manipulation of the trait. We have developed such markers for soybean and used them to genetically characterize seed-coat hardness. Our goals were to identify genes conditioning this trait and to characterize their genetic action. This study illustrates a possible genetic strategy for studying physiological phenomena.

Materials and methods

A population containing great diversity in the hard-seeded trait was chosen for our studies. This was an F_2 population (provided by Dr. W. R. Fehr) generated by crossing a *Glycine max* breeding line (A81-356022 from Iowa State University) and a wild *G. soja* accession (PI 468916). Each of 60 F_2 individuals was grown near Ames/IA and harvested separately. Thirty F_3 seeds from each F_2 individual were scarified and grown in a winter nursery (Isabella, Puerto Rico). The F_4 seeds from each F_2-derived line were bulk harvested and then used in evaluating the seed-hardness trait.

As a measure of seed-coat hardness (permeability), seed germination was evaluated in the absence of scarification. F_4 seeds from each F_2-derived line were placed on wet germination paper in the dark at room temperature for 7 days. The proportion of seeds that germinated was then determined. As a control, all ungerminated seeds were scarified and placed on wet germination paper for an additional 7 days. More than 98% of the ungerminated seeds germinated with scarification. This is consistent with high seed viability and suggests that the seed-coat permeability is a major barrier to germination. The germination assay was replicated twice.

RFLP genotypes of the 60 F_2 plants were determined. DNA was isolated from leaves (Keim et al. 1988) and digested with one of five restriction endonucleases (EcoRI, EcoRV, HindIII, TaqI, or DraI). Digested DNA was then separated by agarose gel electrophoresis (Maniatis et al. 1982) and transferred to nylon membrane (Biotrace RP, Gelman) by using a VacuBlot apparatus (Applied BioNetics). Recombinant DNA clones were isolated from a PstI library (Keim and Shoemaker 1988) and screened for their ability to detect RFLPs in this population (P. Keim, B. W. Diers, R. C. Shoemaker unpublished results). Recombinant DNA was radioactively labelled with ^{32}P (Boehringer-Mannheim random primer kit no. 1004760) and then used for molecular hybridization (Apuya et al. 1988) with DNA from F_2 individuals. F_2 genotypes for five isozymes (malate dehydrogenase, malic enzyme, isocitrate dehydrogenase, acid phosphatase, and diaphorase; Griffin 1986) and three morphological loci (*pb, i,* and dull seed coat; Palmer and Kilen 1987) were determined by screening F_3 progeny.

RFLP genotypes established in the F_2 plants were compared with germination scores in the F_4 progeny of each plant. The families were sorted into genotypic classes (homozygous *G. max*, heterozygous, homozygous *G. soja*) for each of the 72 marker loci. *F*-tests (GLM, Statistical Analysis Systems, Cary/NC) were used to determine if significant differences existed for germination rates among the classes. Loci significant in this analysis were combined into a multivariate linear regression model to determine their combined effects. The significant markers were also tested for epistatic interactions by two-way analysis of variance. The significant interactions and main effects were combined in a multivariate regression model (GLM, SAS) to predict the total variation explained with markers. Genetic linkage among markers was determined by the maximum likelihood method and the computer program MAPMAKER (Lander et al. 1987).

In this study, 72 one-way ANOVAs were performed on the same set of data. By using a significance level of $p<0.05$, there is a great risk of concluding that a marker is linked to a QTL when it is not, whereas choosing a lower level creates the possibility of missing important loci. This dilemma has been resolved by some researchers via a two-step design (Nienhuis et al. 1987). First, significant relationships are observed and then tested in a second set of progeny from the same population. This study constitutes only the observation stage, and markers detecting variation near the $p<0.05$ level will require further confirmation.

Results

In this study, the hard-seed trait was tested on ca. 150 seeds from each of 60 F_2-derived lines. The germination response of the progeny lines was continuous from the extremes of 9% to nearly 100% germination (Fig. 1). The lack of discrete classes is consistent with a polygenically determined trait. Even so, some of the progeny exhibit the extreme parental phenotypes (A81-356022 ca. 100%; PI 468916 ca. 1% germination). This also has been observed on other soybean populations (Kilen and Hartwig 1978), which suggested that a relatively small number of genes (ca. three) was involved in seed-coat hardness. The mean for this segregating population was 53% germination, which is not significantly different from the midparent value.

Populations can be partitioned into genotypic classes defined by genetic markers to identify quantitative trait loci (Edwards et al. 1987). In our study, seed-coat hardnes data were sorted according to marker classes, and *F*-tests were used to determine if significant differences among genotypic classes existed. Seventy-two markers were used, 7 of which described significant effects at the $p<0.05$ level (Table 1). One explanation is that a QTL for hard seededness is located on the same chromosome as

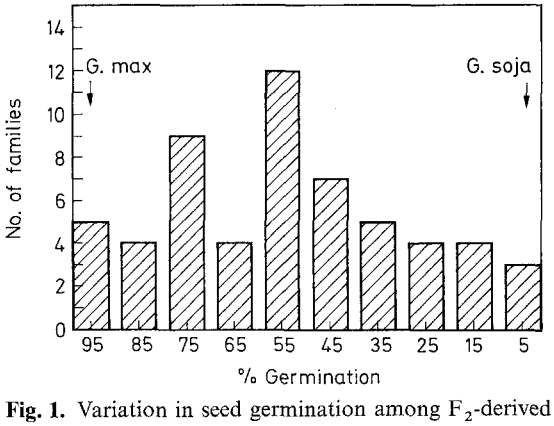

Fig. 1. Variation in seed germination among F_2-derived lines

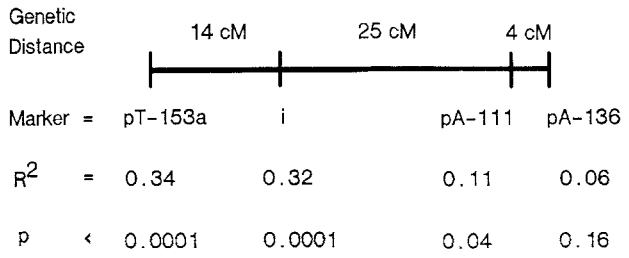

Fig. 2. Genetic linkage of markers predicting variation in seed-coat hardness

Table 1. Markers predicting variation in seed hardness

Marker	One-way ANOVA		Class means[a]		
	R^2	Pr<F	SS	MS	MM
i	0.32	0.0001	0.26	0.52	0.67
pT-153a	0.34	0.0001	0.22	0.53	0.68
pG-17.3a	0.15	0.0140	0.39	0.52	0.39
pK-411	0.13	0.0257	0.49	0.50	0.72
pK-418a	0.12	0.0276	0.35	0.55	0.62
pA-111	0.11	0.0380	0.38	0.56	0.61
pR-22	0.11	0.0488	0.37	0.58	0.58

[a] SS – homozygous *G. soja*; MS – heterozygous; MM – homozygous *G. max*

Table 2. Multiple regression model of hard seed-coat variation

Source	*df*	Sum of squares	Mean squares	*F*-value	Pr>*F*
Model[a]	10	2.05	0.20	6.17	0.0001
Error	46	1.53	0.03		
Total	56	3.57			
$R^2=0.57$					

[a] Model includes all the markers from Table 1 except for pT-153a and pA-111

Table 3. Significant interactions between marker loci

Markers		Pr>F	R^2
i	* pK-418a	0.025	0.13
pK-411	* pT-418a	0.042	0.14
i	* pK-411	0.033	0.11
Combined model[a]		0.0001	0.71

[a] Multivariate model including the main effects from Table 1 and the interactions

the markers and, hence, is genetically linked to the markers. Therefore, the percentage of variation explained by a given marker is a measure of the magnitude of the genetic influence by the QTL. The level of variation explained is modified by the actual genetic linkage between the marker and the QTL; a loose linkage will decrease the variation predicted because recombination dissociates a particular QTL allele from its marker allele.

The amount of genetic variation (R^2) in hard seededness explained by each marker ranged from 11% to 34% (Table 1). The sum of all the markers' R^2 values exceeds 100%. Therefore, different markers may be explaining the same variation in the population. Genetic linkage studies indicated that the *i* locus, pT-153a, and pA-111 are genetically linked (Fig. 2). These three markers seem linked to the same QTL. Because there is evidence (see "Discussion") that the *i* locus itself may be this QTL, the pT-153a and pA-111 markers were not used in subsequent analyses. A fourth marker (pA-136) was also in this linkage group (Fig. 2), but it was not significant in predicting variation in seed-coat hardness. Presumably this was due to the increased map distance between the QTL and pA-136. No other genetic linkage was detected among significant markers.

It is possible that other markers are explaining the same variation, perhaps due to chance colinearity in the data set. Multiple regression models, which combine all the main effects, give a better estimate of the total variation explained than the estimate provided by summing the variation associated with each marker. When the five markers (without pT-153a and pA-111) were analyzed by using multiple linear regression, the total R^2 value was 57% (Table 2), a decrease of 71% from the summed values in Table 1. About 45% of this decrease resulted from the genetically linked markers. The balance may be due to chance colinearity.

Another source of genetic variation in this population are epistatic interactions between genes. All two-way combinations of the significant markers (Table 1) were tested for significant interactions. Three interactions were significant (Table 3). Summed together, the interactions represent 38% of the total variation. The *i* locus is thought to be a regulatory gene that epistatically modifies the action of other genes (Bernard and Weiss 1973). The significant interactions of the *i* marker with markers pK-411 and pK-418a are consistent with the *i* locus itself being the QTL. As with main effects, a particular interaction may be partially explained by other markers or interactions (i.e., colinearity in the data set). A combined model representing both main effects and interactions resulted in an R^2 value of 71% (Table 3). This represents a gain of 14% from the main-effects model (Table 2).

Discussion

Only through the use of molecular markers has it become feasible to systematically dissect the multiple genes that determine many complex traits. Reported here are five independent genomic regions containing putative quantitative trait loci for seed hardness. These regions do not

have equal genetic influence on seed hardness; one region determined more than 30% of the variation (i.e., *i* and pT-153a), whereas another determined only 11% (i.e., pR-22). The presence of "major genes" hidden within the continuous variation of a trait has now been described for other systems as well (Burr et al. 1989). "Minor genes" determine only a small portion of the variation, but create the continuous nature of the trait. The designation of "major" and "minor" genes is not irrevocable, inasmuch as the phenotypic consequence of a particular allele is dependent upon the genetic background (Tanksley and Hewitt 1988) and the environmental situation (Burton 1987). Therefore, what we have described as a "minor" gene could be of major importance in other populations or environments and should not be ignored.

Typically, physiological genetic studies in bacteria, fungi, and *Arabidopsis* have dealt only with single-gene effects (qualitative phenomena) by isolating mutants or by constructing isolines. Great progress has been made in understanding basic metabolism and even developmental processes by using genetics. However, germ plasm in plants that contain single-gene differences is limited, and the physiological consequences of single genes will be oversimplified if the interactions with other genes are not considered. Through the use of segregating populations and RFLP markers, it is now possible to identify different "quantitative" alleles at particular loci and, thus, to study the physiological genetics of polygenic traits.

Our physiological assay is very rudimentary but illustrates the potential of such studies. More sophisticated biochemical and physiological characterization of the seed-coat composition or specific enzymatic activity could be done and then correlated with the known RFLP genotypes. Even without markers, biochemical phenomena can be analyzed for additive and interactive effects in a population segregating for a trait. Then in conjunction with genetic markers, different phenomena could be correlated with QTLs, and the characteristics could be given genetic determinants. Given enough individual QTLs, models can be constructed to explain complex physiological phenomena.

Seed hardness has not been a desired agronomic character and, when soybean was domesticated from wild populations such as *G. soja,* seed hardness would have been selected against. In our population, a cultivar has been crossed with its "putative" wild progenitor (*G. soja*). Genes (QTLs) that we have found to differ between the two parents probably were the objects of selection by early agricultural man. Seed-hardness variation associated with the *i,* pA-111, and pT-153a markers might be determined by segregation of alleles at the *i* locus itself, because seed coat color has been correlated with hardness (Starzing et al. 1982). Although this may be true, it isn't possible to rule out a closely linked gene as the source of this variation. The relatively small number of genes and a partial association with seed color would have facilitated selection against seed hardness during domestication.

A hard seed coat contributes to the viability of stored seeds. This is important to wild soybeans, which may remain dormant for many years before germinating. Increased viability resulting from hard-seeded cultivated soybean becomes important in humid regions such as the southern United States and the tropics, where stored soybeans can deteriorate rapidly. When this occurs, the quality of the commodity decreases. In addition, a harder seed coat is more resistant to soil fungi, which are common in humid areas. A harder seed coat need not be associated with a colored seed (agronomically undesirable), because other genes will contribute to hardness in the absence of the *i* allele (Table 1). In addition, Kilen and Hartwig (1978) have successfully introgressed seed hardness into a soybean breeding line (D76-5679) from *G. soja* in the absence of the *i* allele. Using molecular markers, we have identified five genomic regions associated with seed hardness. These molecular markers could be used by breeders in constructing elite cultivars with varying levels of seed-coat hardness.

Acknowledgements. This work was supported in part by Iowa State Technology Grant no. 480–46–09.

References

Apuya NA, Frazier B, Keim P, Roth JE, Lark KG (1988) Restriction fragment length polymorphisms as genetic markers in soybean, *Glycine max* (L.) Merrill. Theor Appl Genet 75:889–901

Ballard LV (1973) Physical barriers to germination. Seed Sci Technol 1:285–303

Bernard R, Weiss MU (1973) Qualitative genetics. In: Caldwell BE (ed) Soybeans: improvement, product, and uses. American Society of Agronomy, Madison/WI, pp 117–164

Bowman JL, Yanofsky MF, Meyerowitz EM (1989) *Arabidopsis thaliana:* A review. Oxford Surv Plant Mol Cell Biol 5 (in press)

Burr B, Helentjaris T, Tanksley S (eds) (1989) Development and application of molecular markers to problems in plant genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor/NY

Burton J (1987) Quantitative genetics: results relevant to soybean. In: Wilcox JR (ed) Soybeans: improvement, production and uses, 2nd edn. American Society of Agronomy, Madison/WI, pp 211–247

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

Egley GH, Paul RN Jr (1981) Morphological observations on the early imbibition of water by *Sida spinosa* (Malvaceae) seed. Am J Bot 68:1056–1065

Egley GH, Paul RN Jr, Vaugh KC, Duke SO (1983) Role of peroxidase in the development of water-impermeable seed coats in *Sida spinosa* L. Planta 157:224–232

Gots JS, Benson CE (1974) Biochemical genetics of bacteria. Annu Rev Genet 8:79–101

Griffin JD (1986) Genetic and germplasm studies with several isozyme loci in soybean [*Glycine max* (L.) Merr.]. PhD dissertation Iowa State University, Ames, University Microfilms, Ann Arbor/MI, Diss. No. 86–27111

Henry SA, Klig LS, Loewy BS (1984) The genetic regulation and coordination of biosynthetic pathways in amino acid and phospholipid synthesis. Annu Rev Genet 18:207–231

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

Keim P, Olson TC, Shoemaker RC (1988) A rapid protocol for isolating soybean DNA. Soybean Genet Newslett 15:150–152

Kilen TC, Hartwig EE (1978) An inheritance study of impermeable seed in soybeans. Field Crops Res 1:65–70

Lander ES, Green P, Abrahamson J, Barlow A, Daly M, 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

Maniatis T, Fritsch EF, Stambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor/NY

Marbach I, Mayer AM (1974) Permeability of seed coats to water as related to drying conditions and metabolism of plants. Plant Physiol 54:817–820

Marbach I, Mayer AM (1975) Changes in catechol oxidase and permeability to water in seed coats of *Pisum elatius* during seed coat development and maturation. Plant Physiol 56:93–96

Nienhuis J, Helentjaris T, Slocum M, Ruggero B, Schaefer A (1987) Restriction fragment length polymorphism analysis of loci associated with insect resistance in tomato. Crop Sci 27:797–803

Palmer RG, Kilen TC (1987) Qualitative genetics and cytogenetics. In: Wilcox JR (ed) Soybeans: improvement, production, and uses. 2nd edn. American Society of Agronomy, Madison/WI, pp 135–209

Potts HC, Duangpatra J, Hairston WG, Delouche JC (1978) Some influences of hard seededness on soybean quality. Crop Sci 18:221–224

Rolston MP (1978) Water impermeable seed dormancy. Bot Rev 44:365–396

Starzing EK, West SH, Hinson K (1982) An observation on the relationship of soybean seed color to viability maintenance. Seed Sci Technol 10:301–305

Tanksley SD, Hewitt J (1988) Use of molecular markers in breeding for soluble solids content in tomato – a re-examination. Theor Appl Genet 75:811–823

Terzaghi W, Roth EJ, Keim P, Honeycutt EJ, Lark KG (1985) A somatic genetics system for soybean. In: Freeling M (ed) Plant genetics, UCLA Symposia on Molecular and Cellular Biology, New Series 35:771–789

Yaklich RW, Vigil EL, Wergin WP (1984) Scanning electron microscopy of soybean seed coats. Scanning Electron Microsc 2:991–1000