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 F2-derived lines in the F3 generation (F2:3 lines). Each F2:3 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
markers. QTL were found that explained 71% of the total variation for hard seededness in the population. Keim et al. (1990b) 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. (1990a, 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 F2 data from all of the markers were used to map QTL.

Sixty F2 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 F2-derived (F2:3) 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 PsII 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 F2 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 F2:3 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. (1990b) 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 F2:3 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 Gf4 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).