1

9 1

2 6 1 8 9

U·M·I

MICROFILMED 1991

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

U·M·I

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

Order Number 9126189

Restriction fragment length polymorphism mapping of quantitative and qualitative traits in soybean

Diers, Brian William, Ph.D.

Iowa State University, 1991

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

NOTE TO USERS

THE ORIGINAL DOCUMENT RECEIVED BY U.M.I. CONTAINED PAGES WITH PHOTOGRAPHS WHICH MAY NOT REPRODUCE PROPERLY.

THIS REPRODUCTION IS THE BEST AVAILABLE COPY.

Restriction fragment length polymorphism mapping of quantitative and qualitative traits in soybean

by

Brian William Diers

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Department: Agronomy
Major: Plant Breeding and Cytogenetics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa
1991

TABLE OF CONTENTS

GENERAL INTRODUCTION

Soybean [Glycine max (L.) Merr.] is one of the most important oil seed crops in the world (Smith and Huyser, 1987). The expansion of soybean in the USA has been rapid. In the early 1950s, soybean was grown on 5.6 million hectares and in 1980, 28.5 million hectares (Smith and Huyser, 1987). Because of the importance of soybean as a crop, there has been a substantial breeding effort by the private sector, USDA and state agricultural experiment stations to improve soybean (Fehr, 1987).

Although substantial effort has gone into soybean breeding, soybean genetics and cytogenetics are not as advanced as genetics in many other important crops such as maize and tomato. Palmer and Kilen (1987) published a genetic map of soybean which consisted mostly of two and three point linkages. The map contained 35 markers, 17 linkage groups and 420 map units. Soybean cytogenetics also has progressed slowly for the small size and similar morphology of the chromosomes have made the work difficult (Palmer and Kilen, 1987).

The use of restriction fragment length polymorphism (RFLP) markers has spurred rapid progress in the genetics of several crop species (Burr, 1989). In soybean, RFLP markers were developed and used to produce maps both in the public and private sector (Keim et al., 1990b; Tingey et al., 1989). These markers now can be used in soybean breeding and genetics research. The mapping of

quantitative and qualitative traits in soybean with RFLP markers is the purpose of the research presented in this thesis.

LITERATURE REVIEW

Background on RFLP Markers

Genetic markers are useful in breeding and other genetic studies. Until the mid 1980s, the only markers available to researchers were biochemical, such as isozymes, and morphological (Beckmann and Soller, 1983). Although these markers were used extensively, their use was limited by the low number that could be scored in a population. This resulted in the requirement for researchers to combine data from many populations in linkage mapping and also in a poor coverage of the genome in quantitative trait mapping, germplasm surveys and other genetic studies. RFLP markers now offer breeders and geneticists an almost unlimited number of markers that can be used in a given population (Beckmann and Soller, 1983).

RFLP markers are cloned segments of DNA that are used to reveal DNA base pair changes or rearrangements among different genotypes (Beckmann and Soller, 1983). The RFLP genotype of an individual is ascertained first by digesting its DNA with restriction endonucleases. Generally, the digested DNA is separated, based on size, by electrophoresis in an agarose gel, denatured, and blotted onto a membrane. The membrane is hybridized with a cloned DNA segment that has been labelled (Southern, 1975). This DNA clone hybridizes to homologous fragments that are immobilized on a membrane. The sizes of the homologous fragments are determined

by exposing X-ray film with the membrane. If the radiolabelled DNA hybridizes to fragments of differing sizes among individuals, then the cloned segment reveals a polymorphism and can be mapped in a population.

RFLP markers first were discussed as a tool in plant genetics and breeding by Beckmann and Soller (1983). Since that time, the use of RFLP markers has resulted in an unprecedented increase in genetic mapping in many crop species (Burr, 1989). By 1989, molecular maps of eight plant species were completed (Burr, 1989). In contrast, before the use of RFLP markers, classical genetic maps were formed through many decades of effort.

Soybean RFLP Mapping

Apuya et al. (1988) investigated the use of RFLPs as genetic markers in soybean. They reported the level of RFLP diversity in G. max was low with 20% of the random clones revealing polymorphisms between the widely distant cultivars 'Minsoy' (PI 27890) and 'Noir I' (PI 290136). They observed that a high proportion of the polymorphisms were caused by genome rearrangements. Apuya et al. (1988) also screened five soybean cultivars for RFLP polymorphisms and found most markers having only two alleles. This is in contrast with maize where more than two alleles for each locus are common (Burr et al., 1983).

Keim et al. (1989) surveyed 58 soybean accessions from the genus Glycine, subgenus Soja with 17 RFLP markers. They found very low levels of diversity among North American cultivars. Seven

of these cultivars were identical for all 17 markers. Greater diversity was observed between the G. max cultivars and G. gracilis Skvortz. or G. soja Sieb. and Zucc.accessions. Like Apuya et al. (1988) multiple alleles at a locus were infrequent among all accessions surveyed. Only two of the 17 markers revealed loci with greater than two alleles. The authors suggested using an interspecific population for RFLP mapping in soybean to take advantage of the higher level of diversity among species than within species.

Keim et al. (1990b) reported a soybean RFLP map that was made using a population formed from an interspecific cross between G. max and G. soja. This map was made by Iowa State University and the USDA-ARS. The map contained 150 markers, 26 linkage groups and covered 1,200 centiMorgans (cM). Forty percent of the random clones tested from their Pst1 library revealed a polymorphism between the two parents which was much higher than within G. max. Because the soybean genome consists of 20 chromosomes, (Hymowitz and Singh, 1987), extra linkage groups exist on their map and should later be joined after more markers are added. More markers are being added to this map and presently it contains over 300 markers.

A soybean RFLP map also was made in the private sector. Researchers at Du Pont de Nemours and Co. made a RFLP map in a population formed from a cross between G. max and G. soja (Tingey et al., 1989). Only 10% of the random clones tested revealed polymorphisms between the two parents which is much lower than the 40% found by Keim et al. (1990b) between their G. max and G.

soja parents. The Du Pont map contained about 550 markers covering 2,700 cM in 23 linkage groups (Rafalski and Tingey, 1990).

RFLP data can give insight into the origin of the soybean genome. Soybean is thought to have been a tetraploid that diploidized (Hymowitz and Singh, 1987). Essentially all clones used as RFLP markers revealed two or more fragments by Southern hybridizations to genomic DNA that was restriction digested. The multiple fragments were expected for an organism with polyploid origins for it is indicative of duplicate copies of a genome (Keim et al., 1990b). Studying the clones that map more than one locus also gives information on the structure of the soybean genome. Keim et al. (1990b) found that twenty-three of the clones used as markers mapped to more than one locus. If few major DNA rearrangements have occurred since the doubling of the genome, two clones that each map two loci and are linked in one linkage group would be expected to be linked in a second group because both groups would be on homoeologous chromosomes. This was not found, for most of the second loci were unlinked. It indicates that if soybean has tetraploid origins, much scrambling of the genome has occurred since the tetraploidization (Keim et al, 1990b). This scrambling is in contrast to findings in maize by Helentjaris et al. (1988). They could identify duplicated linkage groups using linked clones that mapped two loci which they attributed to the existence of ancient homoeologous chromosomes.

Qualitative Trait Loci Mapping with RFLP Markers

The integration of qualitative trait loci (QLTL) onto RFLP maps would increase the usefulness of the linkage maps. The QLTL could be mapped by two methods. The traditional method would involve screening segregating populations for linkage between markers and QLTL. A F_2 population size of 34 would be required to detect a linkage of 20 recombination units or less at a 0.95 probability level between a codominant marker and a dominant trait (Hanson, 1959). The traditional approach therefore would require screening a large number of markers against a population containing many individuals to have a high probability of detecting linkage.

An alternative approach for mapping QLTL is the use of near-isogenic lines (NILs). A NIL is formed by repeatedly backcrossing a QLTL from a donor parent into the background of a recurrent parent. After repeated backcrossing and selection for a QLTL, most of the genome of the donor parent will be eliminated with the exception of a linkage block surrounding the gene of interest. Markers that are likely linked to the QLTL could be found by surveying markers to find those that are polymorphic between the recurrent parent and the NIL. These polymorphic markers would have a high probability of being linked to the QLTL and could be tested for linkage in a population using standard linkage tests. A large stock of NILs currently exists in soybean (Bernard, 1976) and these could be used for mapping QLTL.

According to formulas summarized by Muehlbauer et al. (1988), a soybean NIL formed through six backcrosses would have,

on average, 2.1% of its genome originating from the donor parent. Of the donor parent's contribution to the NIL, 65% would be in a linkage block surrounding the QLTL backcrossed into the NIL. The calculations assume a genome of 20 chromosomes with each being 100 cM in length and no selection during backcrossing other than for the QLTL being backcrossed into the NIL.

A high proportion of the markers used to screen NILs must be polymorphic between the recurrent parent and the donor parent for the isoline method to be successful in mapping QLTL. If a marker maps within a linkage block that originates from the donor parent but the marker is not polymorphic between the donor and recurrent parents, the linkage block would not be detected with the marker. Muehlbauer et al. (1989) surveyed 63 donor parents of soybean NILs and their recurrent parents with 12 isozyme loci. They observed that 34% of isozyme markers were polymorphic between the donor parents and their recurrent parents. The authors stated that this level is sufficient for the use of NILs as a mapping tool. Muehlbauer et al. (1990) tested 116 donor parents of NILs with 15 RFLP markers and found the level of polymorphism between the donor and recurrent parents to be 34%. The authors found this level strikingly similar to the level found with isozymes and it also confirmed the usefulness of NIL for mapping QLTL.

NILs have been used to map QLTL. Young et al. (1988) used NILs in tomato (Lycopersicon esculentum) to obtain two RFLP markers tightly linked with the Tm-2a gene, which confers

resistance to tobacco mosaic virus. These two tightly linked markers were the only polymorphisms found between the NILs and the recurrent parent. Gardiner et al. (1989) used NILs to map the Ht locus in maize (Zea mays L.). Ht confers resistance to Helminthosporium turcicum. Muehlbauer et al. (1989) found polymorphisms between five NILs and their recurrent parents using isozymes in soybean. Two of the linkages were confirmed and two refuted using F_2 segregation data. The fifth presumptive linkage was not tested. The two loci mapped were the ln locus conferring ovate or lanceolate leaf shape and the dt2 locus conferring semideterminate or indeterminate stem-growth habits (Muehlbauer et al., 1989). Muehlbauer et al. (1990) also found 15 polymorphisms between NILs and their recurrent parents using RFLP markers in soybean. Two of the presumptive linkages were confirmed using F_2 segregation analysis. The two genes mapped were r-m, that confers black stripes on brown seed, and lfl, that confers pentafoliolate or trifoliolate leaf number (Muehlbauer et al., 1990).

Quantitative Trait Loci Mapping with RFLP Markers

Many traits that are of the greatest importance to plant breeders are quantitatively inherited. For this reason, there has been much interest in mapping quantitative trait loci (QNTL) in plants with RFLP markers. Before the introduction of RFLP markers in plants, researchers using morphological and biochemical markers mapped some QNTL (Tanksley et al., 1982; Edwards et al. 1987). These researchers were, however, limited by the low number of

markers that were segregating in their populations (generally fewer than 20). The low number of markers resulted in a lack of adequate coverage of the genome. This resulted in many important QNTL being missed or their effects underestimated because of loose or no linkage between markers and QNTL.

It is now possible to more fully cover the genome of plants in a single population using RFLP markers. The use of RFLP markers to map QNTL first was reported by Osborn et al. (1987). They found one location in the genome that was associated with soluble solid content in tomato (Lycopersicon esculentum). Soluble solid content also was studied in tomato by Tanksley and Hewitt (1988). They found three chromosome segments associated with soluble solid content. The effect on soluble solids associated with one of the segments was population specific.

Several more mapping studies of QNTL with RFLP markers recently have been published. Nienhuis et al. (1987) mapped insect resistance in tomato with RFLP markers. They found four loci that in a multivariate model explained 38% of the phenotypic variation for a factor involved in insect resistance. Martin et al. (1989) found three RFLP markers significantly associated with water use efficiency in tomato. Grant et al. (1989) mapped QNTL for maturity related traits in maize. They observed that some loci explained 10-20% of the total variation for these traits. Bubeck et al. (1990) mapped gray leaf spot (caused by Cercospora zeae-maydis) in maize with RFLPs. Webb et

al. (1990) obtained associations between RFLP markers and seed fatty acids, seed oil, and embryo weight in Cuphea lanceolata.

Both isozymes and RFLPs have been used to identify QNTL in soybean. Graef et al. (1989) and Suarez (1989) studied the association between isozyme markers and quantitative traits in two populations formed through backcrossing G. max. to G. soja. Both found significant associations between vegetative traits and isozyme loci. Suarez also found several associations between specific isozyme markers and protein, oil and relative 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 x G. soja single-cross population with 70 RFLP markers. QNTL were found that explained 71% of the total variation for hard seededness in the population using a multivariate model that included five significant markers.. Keim et al. (1990b) continued mapping in the same population and found significant QNTL for maturity and morphological traits by using 150 RFLP markers. For several traits, they found markers that individually explained more than 20% of the total variation.

QNTL have been traditionally mapped using one-way ANOVAs for each marker trait combination (Sax, 1923; Tanksley et al., 1982; Edwards et al., 1987). For each marker, the mean of the individuals in each genotypic class was tested to determine if they were

significantly different. If the means were significantly different, it indicated a significant association between the marker and gene(s) controlling the QNTL.

Lander and Botstein (1989) introduced interval mapping, an alternative method of mapping QNTL. The interval mapping method predicts the phenotypic effect for a putative QNTL at any location in the genome. This method allows researchers to estimate the location and effect of the QNTL more accurately than traditional methods. The use of the interval mapping method was demonstrated by Paterson et al. (1988). They found several QNTL that were significantly associated with fruit mass, concentration of soluble solids, and fruit pH in tomato.

Conclusions

The technology and methods have been developed that makes efficient RFLP mapping in soybean a reality. RFLP markers and maps are useful to the soybean geneticist, but themselves have limited application to soybean breeders. The linkage of RFLP markers to important traits could result in greater application of these markers in breeding. These linkages could allow breeders to dissect quantitative traits to help formulate breeding strategies, or to select individuals based on marker genotypes. In this thesis, I endeavor to use RFLP markers to obtain information on genes in soybean that could be useful to plant breeders.

Explanation of the Dissertation Format

This dissertation contains three sections. Section I reports the continuation of QNTL mapping that was initiated by Keim et al. (1990a,b). In this study, QNTL for seed size, total protein, and oil content and fatty acid content were identified in the G. max by G. soja mapping population (Keim et al. 1990b). Section II described the mapping of phytophthora resistance. Section III discusses a glycinin storage protein variant that was identified in the G. max by G. soja mapping population (Keim et al., 1990b).

Each section is a complete manuscript. Sections II and III have been or will be submitted for publication as is. Section I however will require revision or an extra year of data before it can be submitted. References cited in the general introduction and literature review are listed in "Additional References Cited" following the General Conclusion.

SECTION I: RFLP ANALYSIS OF SOYBEAN SEED SIZE, PROTEIN, OIL, AND FATTY ACID CONTENT

ABSTRACT

The objective of this study was to identify quantitative trait loci (QTL) in soybean [Glycine max (L.) Merr.] for seed size, total protein and oil content, and content of the fatty acids palmitate, stearate, oleate, linoleate and linolenate. Two populations were measured in this experiment. Both populations were formed from a cross between a G. max experimental line (A81-356022) and a G. soja Sieb. and Zucc. plant introduction (PI 468916). In population 1, all traits 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. In population 2, seed size, and total protein and oil content were measured on seed harvested from 58 F_2 plants. These plants were genotyped with three RFLP markers that are on the two linkage groups most highly associated with protein and oil in population 1. Significant ($P<0.01$) associations were found between the segregation of markers and all traits in population 1. Segregation of individual markers explained up to 43% of the total variation for specific traits. Many of the markers significantly associated with variation for seed traits were clustered on linkage groups, 'A', 'B' and 'K'. 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. G. max alleles at significant loci for seed size were all associated

with larger seed size than G. soja alleles. No consistent relationship was detected between the species origin of alleles and the content of any fatty acid. For each fatty acid, G. max alleles at significant loci were associated with both greater and lesser levels of the fatty acid than G. soja alleles. The results from population 2 showed the associations between markers and protein and oil were consistant between populations for one linkage group but not the other. The most likely cause for this discrepancy is genotype by environoment interactions, although type I errors can not be discounted.

Additional index 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 may cause protein and oil quality to become major breeding objectives in the future. Seed size also has received emphasis because of its importance in some specialty markets.

The oil of soybean seed is composed of five major fatty acids. Palmitate and stearate are 16- and 18-carbon saturated fatty acids, respectively, and oleate, linoleate and linolenate are unsaturated 18-carbon fatty acids with one, two and three double bonds, respectively. Breeders have worked to reduce the level of linolenate in soybean because it is inversely associated with stability of the oil (Wilson 1987).

Glycine soja, a wild relative and putative ancestor of Glycine max, offers a source of genes that may be useful in the improvement of cultivated soybean. Both G. max and G. soja share the same subgenus and are mostly interfertile (Hymowitz and Singh 1987). Breeders have used G. soja as a source of small seededness and high protein content (Erickson et al. 1981). The use of G. soja in the improvement of soybean is difficult because it has many undesirable characteristics. Ertl and Fehr (1985) observed that populations derived from G. soja x G. max crosses required two backcross generations to recover segregates phenotypically similar to the G. max recurrent parent.

Soybean seed size, protein, oil and fatty acid content are quantitatively inherited (Weber 1950; Burton 1985; Wilcox 1985; Howell et al. 1972). Breeders have been successful in the manipulation of 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 QTL mapping 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 x G. soja derived populations. Graef et al. (1989) and Suarez (1989) studied the association between isozyme markers and quantitative traits in two G. soja x 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 x 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 size, protein and oil content, and fatty acid content in the same population used by Keim et al. (1990a,b) and in a second population.

MATERIALS AND METHODS

The study was conducted with two populations, each produced from a cross between the G. max experimental line A81-356022 and the G. soja accession PI 468916. Population 1 was used to develop the RFLP map and the F_2 data from all of the markers were used to map QTL. Population 2 was used as an independent test for some of the most important QTL found in population 1.

Sixty F_2 plants from population 1 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.

Population 2 consisted of 58 F_2 individuals that were grown from February to June 1990 in a greenhouse in Ames, IA. The plants were grown in pots 25 cm in diameter. They were exposed to artificial lighting set at a 16-hour day length from germination to 45 days post germination and then a 14-hour day length until maturity. As with population 1, leaf samples were taken from each F_2 plant for DNA extractions, and at maturity, each plant was harvested and threshed separately.

Linkage Mapping

A total of 252 loci were scored using population 1 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 sixteen loci scored using recombinant DNA clones obtained from other labs, five isozyme markers (Rennie et al., 1989), 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 by 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 for population 1 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.

The traits seed size, protein and oil content and fatty acid content were determined for F_4 seed harvested from each plot for population 1. The traits seed size and protein and oil content were measured on F_3 seed harvested from each F_2 individual for population 2. Seed size was

determined by weighing a 50-seed sample. Seed protein and oil content was measured from a 5- to 7-g ground sample at the USDA Northern Regional Research Center at Peoria, IL, by using a Pacific-Scientific NIR grain analyzer.

Fatty acid content was measured on a 10-seed sample by gas chromatography. The analyses were done by Dr. E. G. Hammond and D. N. Duvick in the Department of Food Technology at Iowa State University. The seeds were crushed between two aluminum plates by using a hydraulic press and a pressure of 1500-3000 Kg cm^{-1}. Five-seed samples were crushed in grooves 8 mm wide and 1.6 mm deep. The sample then was transferred to a test tube and extracted approximately 18 hr with sufficient hexane to cover the seed. Approximately 0.1 ml of the hexane extract was transferred to a 1.5-ml autosampler vial and reacted with 0.5 ml of 1 M sodium methoxide solution in methanol for 30 min at 40^{o} C with gentle mixing for 10 min. Next, 0.8 ml of water was added, and after 3-5 min, the floating oil was dissolved in 0.5 ml hexane. The hexane layer was transferred to a clean vial and diluted to about 1.1 ml with additional hexane. The samples were analyzed in a Hewlett-Packard (Avondale, PA) 5890 gas chromatograph fitted with flame detectors and 15-M Durabond-23 capillary columns (J&W Scientific, Deerfield, IL) 0.25-mm i.d. and 0.25-u film thickness at 200^{o} C. Percentages were calculated from electronically integrated peak areas corrected by factors based on the number of C-H bonds. The reliability of these correction factors was verified daily by injecting a standard of known composition.

Seed trait data for population 1 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 markers 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 markers.

The individuals for population 2 were genotyped with three markers that were significantly associated to the seed traits in population 1. The markers were pSAC-7a and pb in linkage group 'A' and pK-11a in linkage group 'K'. pSAC-7a was a actin gene probe that was kindly provided by Dr. Richard Meagher from the University of Georgia, pb was a morphological gene that confers blunt-sharp pubescence tip and pK-11a was a recombinant soybean clone developed for RFLP mapping. The associations between the traits and markers

were tested using the same analysis procedures described for population 1.

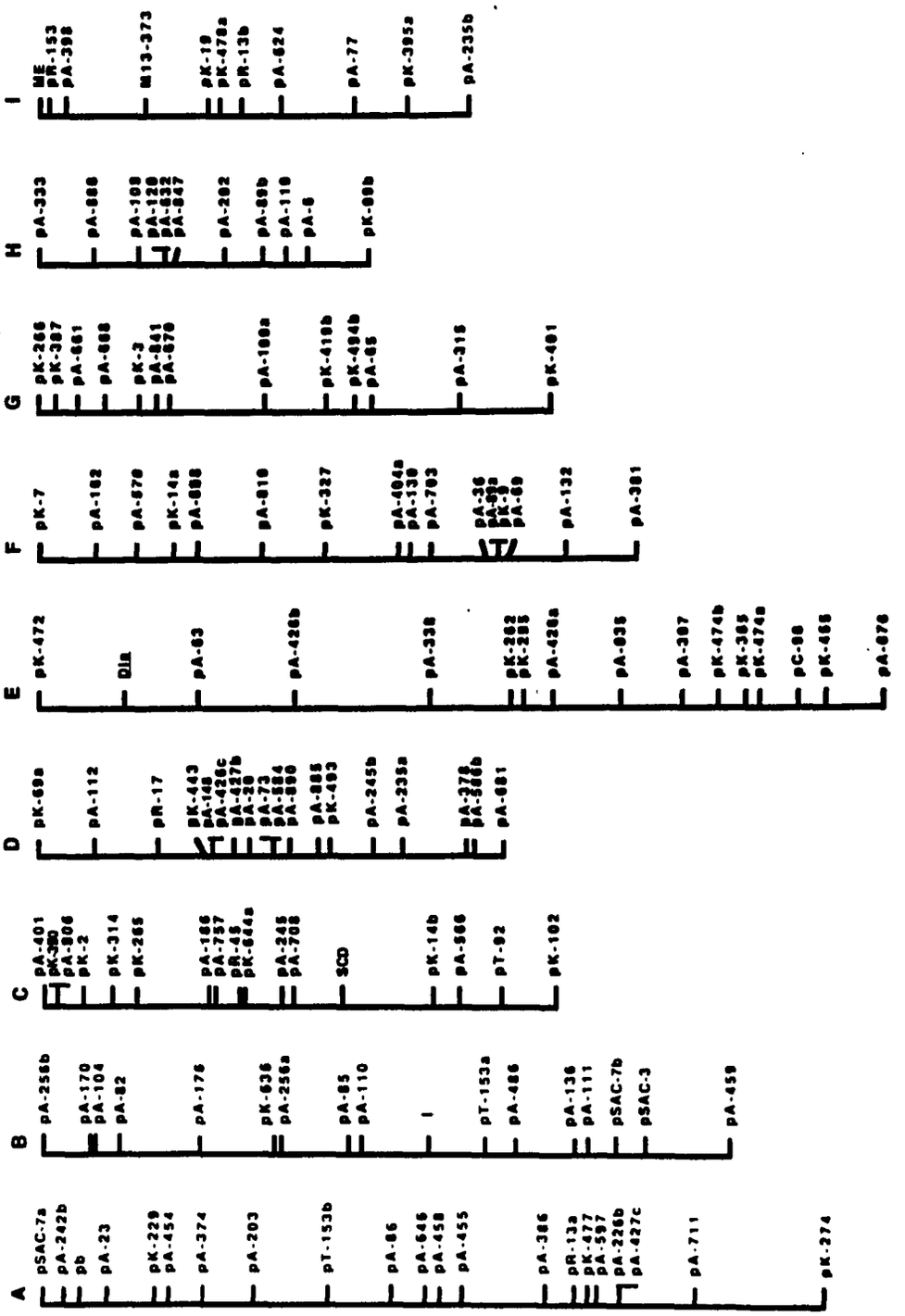
RESULTS

The RFLP data were used to construct a RFLP linkage map of soybean (Figure 1). The map contains 252 markers, 31 linkage groups and 2,147 centiMorgans. The map contains linkage groups that 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 all measured traits based on analysis of seed from the 1988 field trial (Table 1). The G. soja parent had a greater protein level and a lesser oil level than the G. max parent. The parents contained similar levels of individual fatty acids except for oleate and linolenate. The G. max parent contained 44 g kg^{-1} more oleate and 62 g kg^{-1} less linolenate than the G. soja parent. The parents were greatly different for seed size with the G. max parent having seed that were 11-times larger than the G. soja parent.

Significant genetic variation was present among the $F_{2:3}$ lines for all traits in population 1. 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 the seed traits at $P < 0.01$ are given in Table 2. Some markers that were associated with significant

Figure 1. Soybean RFLP map. The names of the linkage groups are listed at the top of each group. The markers labelled pA and pK are RFLP markers developed 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. Richard 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
I
pT-153a
pA-486
pA-136
pA-111
pSAC-7b
pSAC-3
pA-458
C
pA-401
pK-380
pA-806
pK-2
pK-314
pK-285
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-38
pA-73
pA-884
pA-880
pA-885
pK-493
pA-245b
pA-235a
pA-378
pA-388b
pA-681
E
pK-472
Dia
pA-83
pA-426b
pA-338
pK-282
pK-285
pA-426a
pA-836
pA-387
pK-474b
pK-385
pK-474a
pC-88
pK-466
pA-676
F
pK-7
pA-182
pA-870
pK-14a
pA-888
pA-810
pK-327
pA-404a
pA-130
pA-703
pA-36
pA-88
pK-5
pA-80
pA-132
pA-381
G
pK-266
pK-387
pA-661
pA-868
pK-3
pA-841
pA-870
pA-188a
pK-418b
pK-484b
pA-85
pA-315
pK-481
H
pA-333
pA-886
pA-108
pA-120
pA-833
pA-847
pA-282
pA-89b
pA-110
pA-6
pK-88b
I
ME
pR-153
pA-398
M13-373
pK-19
pK-478a
pR-13b
pA-624
pA-77
pK-385a
pA-235b

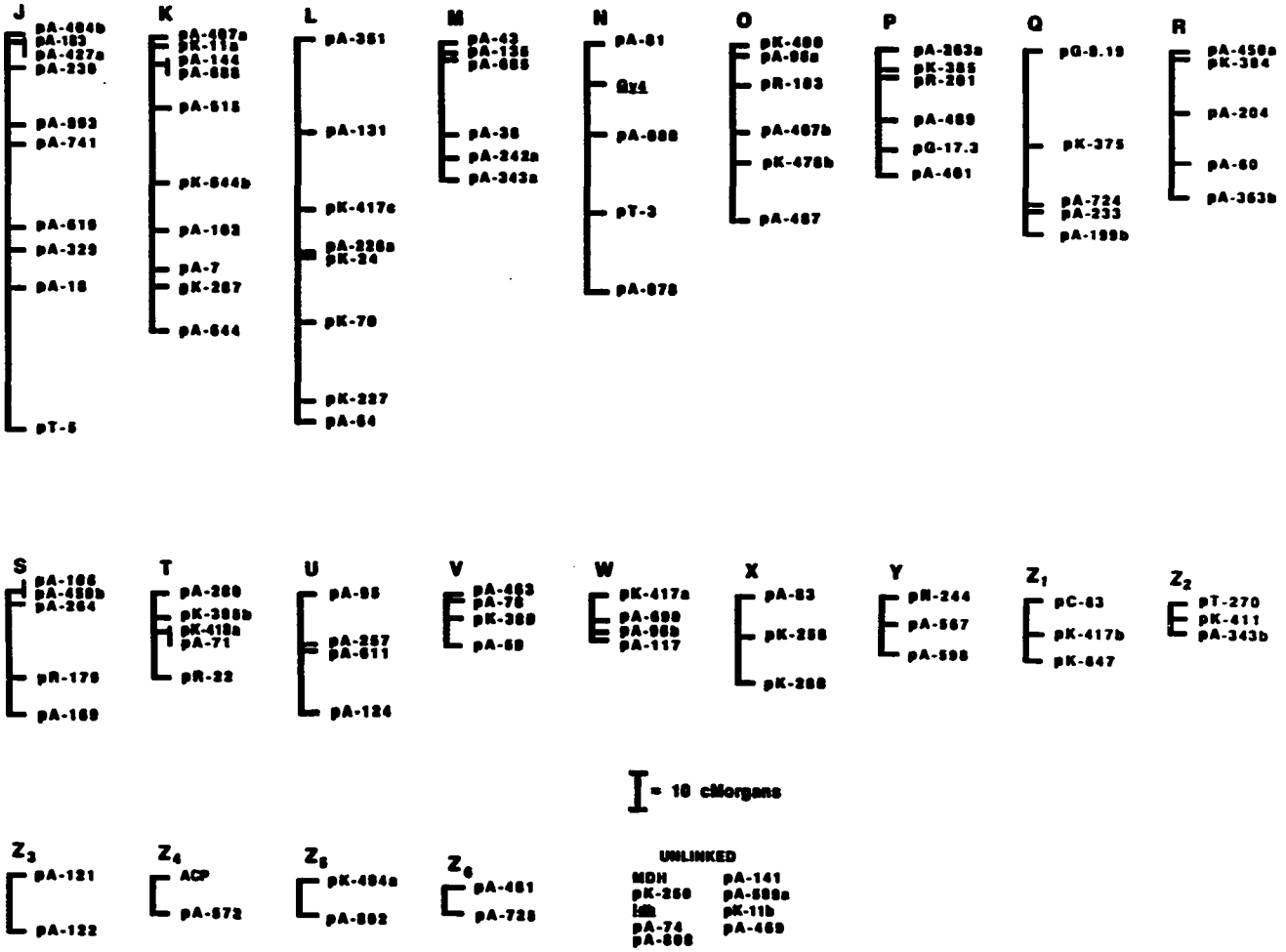


Figure 1. (Continued)

Table 1. Mean performance of the G. max and G. soja parents, and the broad-sense heritabilities of the population

Trait	Mean Performance G. max A81-356022	Mean Performance G. soja PI468.916	Heritability
	—g (kg seed)$^{-1}$—		
Protein	420†	471	0.74
Total Oil	198	101	0.92
	—g (kg oil)$^{-1}$—		
Palmitate	102	112	0.85
Stearate	52	41	0.86
Oleate	217	173	0.86
Linoleate	544	528	0.88
Linolenate	84	146	0.86
	—mg seed^{-1}—		
Seed size	176	16	0.84

†Values for all traits are significantly different between parents ($P<0.001$).

Table 2. Markers significantly (P<0.01) associated with variation for seed traits in population 1

Marker	R^2	P>F	Means of genotypic classes† MM	SM	SS	Linkage Group
			—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‡	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§	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¶	0.27	0.0001	152	144	139	A
pK-11A	0.27	0.0002	155	144	140	K

†MM designates homozygous G. max class, SM heterozygous class and SS homozygous G. soja.

‡Actin gene probe provided by Dr. Richard Meager (University of Georgia).

§G. soja alleles dominant to G. max alleles.

¶Morphological marker blunt-sharp pubescence tip (Palmer and Kilen 1987).

Table 2. (Continued)

Marker	R^2	P>F	Means of genotypic classes MM	SM	SS	Linkage Group
			—g (kg seed)$^{-1}$—			
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
			—g (kg oil)$^{-1}$—			
Palmitate						
pA-343a	0.24	0.001	112	114	109	M
pA-18	0.19	0.005	115	111	111	J
pK-375	0.18	0.005	111	111	115	Q
pK-11b#	0.12	0.01	111		115	Unlinked
Stearate						
pA-233	0.19	0.005	42	44	46	Q
pA-203	0.18	0.008	44	44	41	A
pA-242b	0.16	0.008	44	44	41	A
Oleate						
pA-82	0.28	0.0001	176	189	199	B
pA-104	0.26	0.0003	175	188	198	B
pA-170	0.23	0.001	177	188	198	B
pb	0.21	0.001	192	191	175	A
pA-242b	0.20	0.002	192	191	176	A

#G. max alleles dominant to G. soja alleles.

Table 2. (Continued)

Marker	R^2	P>F	Means of genotypic classes MM	SM	SS	Linkage Group
			—g (kg oil)$^{-1}$—			
pA-619	0.19	0.004	179	193	187	J
pA-454	0.16	0.008	193	190	178	A
pA-203	0.16	0.01	190	191	177	A
pR-13a	0.16	0.008	192	189	174	A
pT-153b#	0.12	0.008	189		174	A
Linoleate						
pA-82	0.38	0.0001	558	546	534	B
pA-104	0.33	0.0001	559	547	536	B
pA-170	0.30	0.0001	557	547	535	B
pA-242b	0.21	0.002	545	543	557	A
pA-118	0.20	0.006	537	549	550	H
pb	0.20	0.002	544	543	557	A
pA-256b	0.18	0.008	556	546	541	B
pA-176	0.17	0.008	556	546	538	B
pT-153b#	0.15	0.004	545		560	A
Linolenate						
pSAC-7a	0.31	0.0005	104	110	116	A
pA-23	0.26	0.0005	106	111	115	A
pA-242b	0.23	0.0008	106	111	115	A
pA-203	0.22	0.002	108	109	116	A
pA-454	0.22	0.001	106	110	115	A
pA-65	0.20	0.003	111	113	105	G
pK-229	0.20	0.003	107	110	116	A

Table 2. (Continued)

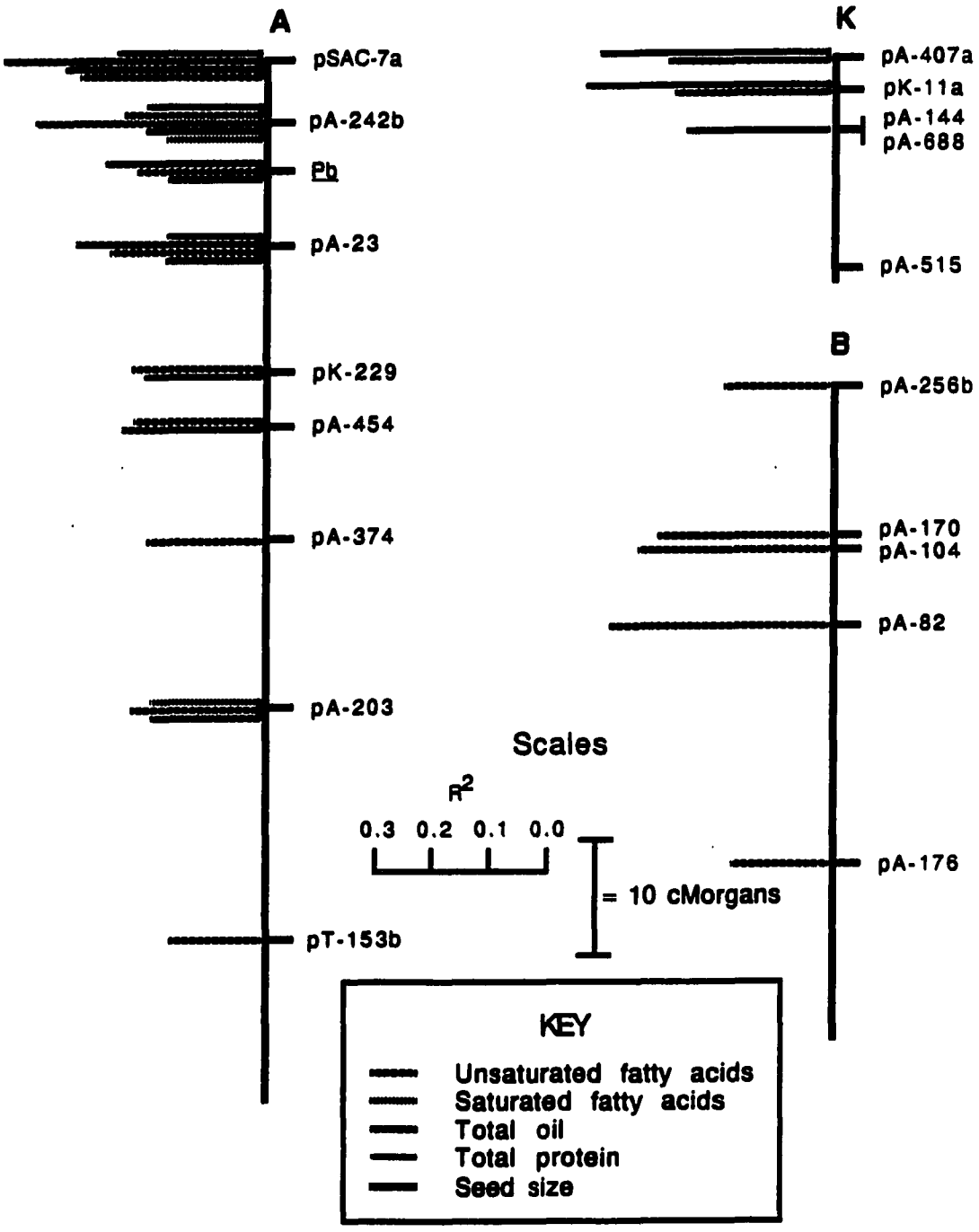
Marker	R^2	P>F	Means of genotypic classes[a] MM	SM	SS	Linkage Group
			—g (kg oil)$^{-1}$—			
pb	0.19	0.003	106	111	115	A
pA-374	0.19	0.003	108	108	115	A
			—mg seed^{-1}—			
Seed Size						
pSAC-7a	0.33	0.0003	48	43	39	A
pK-472	0.21	0.002	48	42	40	E
pA-242b	0.19	0.004	48	42	40	A
pA-23	0.16	0.009	48	42	40	A
pA-245A	0.16	0.01	48	42	46	C
pb	0.15	0.009	48	42	40	A

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 2, Figure 2). The 'A' linkage group in particular was associated with significant variation for all traits measured, except palmitate.

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 less 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 (Figure 2). Alleles significantly associated with increased levels of one trait, protein or oil, generally were associated with lower levels of the other trait (Table 2).

At loci significant for fatty acid content, marker alleles from both the G. max and G. soja parent were associated with both high and low levels of each fatty acid (Table 2). The significant markers for fatty acids were mostly on the 'A', and 'B' linkage groups (Figure 2). The G. max alleles at significant loci in the 'A' linkage group were associated with greater content of stearate and oleate or lesser content of linoleate and linolenate than alleles from G. soja. Effects associated with G. max alleles in the 'B' group were somewhat opposite the effects of the G. max alleles in the 'A' group. The G max alleles in the 'B' group were associated with lesser content of oleate and greater content of linoleate

Figure 2. Linkage groups with the greatest association to QTL. Marker names are listed to the right of each linkage group. The thick horizontal bars represent associations to QTL


A
pSAC-7a
pA-242b
Pb
pA-23
pK-229
pA-454
pA-374
pA-203
pT-153b
K
pA-407a
pK-11a
pA-144
pA-688
pA-515
B
pA-256b
pA-170
pA-104
pA-82
pA-176
Scales
R2
0.3 0.2 0.1 0.0
= 10 cMorgans
KEY
Unsaturated fatty acids
Saturated fatty acids
Total oil
Total protein
Seed size

than G. soja alleles. The effect of markers in the 'B' group was not significant ($P<0.01$) for linolenate. There was a significant effect from a marker on the 'G' group for linolenate, with G. max allele associated with greater linolenate than G. soja alleles.

An 11-fold difference in seed size was observed between the G. max and G. soja parent (Table 1). All significant G. max alleles were associated with larger seed size than G. soja alleles (Table 2). Four of these markers were on linkage group 'A' (Figure 2).

The loci pK-11a in linkage group 'K' and pSAC-7a and pb in linkage group 'A' were tested for associations with protein, oil and seed size in population 2. The association of marker pK-11a with protein and oil content were similar between populations 1 and 2 especially for the difference between the means of the homozygous classes (Tables 2 and 3). In population 1, the difference between the mean of the individuals that are homozygous G. max and G. soja is 2.4 g(kg seed)$^{-1}$ for protein and 1.5 g(kg seed)$^{-1}$ for oil. In population 2, the difference is 2.2 and 1.8 g(kg seed)$^{-1}$ for protein and oil respectively.

The association of pSAC-7a and pb with seed size, protein and oil content were not consistent between populations. Very large associations were found between these markers and the three traits in population 1, but no significant associations were found in 2.

Table 3. Markers tested for association with variation for seed traits in population 2

Marker	R^2	P>F	Means of genotypic classes† MM	SM	SS	Linkage Group
			—g (kg seed)$^{-1}$—			
Protein						
pK-11a	0.23	0.001	425	443	447	A
pSAC-7a‡	0.04	0.433	439	445	436	K
pb¶	0.01	0.669	436	441	438	A
Total Oil						
pK-11a	0.11	0.04	189	179	171	A
pSAC-7a	0.13	0.06	173	178	190	K
pb	0.00	0.96	178	180	179	A
			—mg seed^{-1}—			
Seed Size						
pK-11a	0.01	0.76	41	43	41	A
pSAC-7a	0.01	0.83	40	42	41	K
pb	0.06	0.17	46	40	41	A

†MM designates homozygous G. max class, SM heterozygous class and SS homozygous G. soja.

‡Actin gene probe provided by Dr. Richard Meager (University of Georgia).

¶Morphological marker blunt-sharp pubescence tip (Palmer and Kilen 1987).

DISCUSSION

Regions of the genome that explain large amounts of variation for seed size, protein, oil and fatty acid content were identified with population 1. 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. Loci segregating for genes with large effects in proportion to the total phenotypic variation could be expected in a diverse, interspecific population as was used in this study.

There is a good probability that markers were erroneously declared significant in population 1 which would have resulted in Type I errors. With the probability cutoff of 0.01 that was used in this experiment, the probability 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 lowers to 0.22 for a probability level of 0.001 and to 0.025 for a probability level of 0.0001. Although the probability 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 2).

The markers SAC-7a and pb from group 'A' and pK-11a from group 'K' were tested in population 2 to determine if trait-marker associations were consistent between populations. The associations were consistent with pK-11a but not for either pSAC-7a or pb. With

pK-11a, the differences among the means for the genotypic classes were similar for both populations (Tables 2 and 3). However, in population 2 the R^2 values were smaller and the P>F larger than in population 1. A reason for these differences is that in population 2 the traits were measured for individual plants which would give less accurate estimates of the genotypic values of the F_2 individuals than in population 1 where progeny testing was done.

There are a few possibilities to explain the inconsistency between populations for the associations of the seed traits with pSAC-7a and pb that are both in group 'A'. The most plausible explanation is the effect of genotype by environment interactions. The seed analyzed for the traits came from very different environments for each population. The seed analyzed in population 1 was harvested from a replicated field trial and the seed for population 2 was harvested from greenhouse-grown plants. The expression of the seed traits could be differentially effected by these environments.

Another explanation is that the significant effects associated with group 'A' in population 1 were the result of Type I errors. This would mean that there are no actual QTLs for seed size and protein and oil content in the upper part of linkage group 'A'. This seems somewhat unlikely because of the very high level of significance associated with these markers.

Another explanation for the inconsistency is the traits were measured in population 2 on single plants and in population 1 in a replicated trial. As described previously, the genotypic values of the

F_{2}s would be less accurately estimated using the single plant data than data from the replicated test. The lower accuracy would make finding significant marker trait associations more difficult in population 2 than 1.

Seed protein and oil content are generally negatively correlated in soybean breeding populations (Burton 1985), and in population 1, the phenotypic correlation was -0.73. The RFLP analysis in this population was consistent with this negative correlation. The analysis showed that regions of the genome associated with high levels of one trait were generally associated with low levels of the other (Table 2). If one considers linkage groups 'A' and 'K', which had the greatest association with protein and oil content, the G. max alleles at significant loci were associated with high levels of oil and low levels of protein.

The extreme difference in seed size between the two parents of the population used in this study may have a large impact on the associations between the markers and other seed traits analyzed. Presumably, this extreme difference could result in the other association being an artifact of seed size. Although the analyses used in this study do not give a direct answer to this concern, there is some evidence that the impact may not be very large. The evidence is that although the parents ranged in seed size from 176 to 16 mg $seed^{-1}$ in the field trial, genes that had a large effect on seed size were not detected in the population. This was shown because no markers with large effects compared with the difference between the parents were found. The variation in seed size among the progeny also was relatively

low with a standard deviation of 6.48 mg sd^{-1} in population 1. This low standard deviation is characteristic for a trait controlled by many genes with each having a small effect. There were, however, significant ($P>0.01$) phenotypic correlations between seed size and seed content of total oil (0.58), stearate (0.35), oleate (0.41), and linolenate (-0.62) in population 1. These correlations fit with the associations found in linkage group 'A', with G. max alleles associated with larger seed, greater total oil content and lesser linolenate content. This indicates that the correlations are at least partially caused by linkage between genes for seed size and the other traits or by pleiotropy of genes that affect seed size and the other correlated traits.

Three linkage groups were associated with major effects for fatty acid content in population 1. It is likely that the significant loci in groups 'A' and 'B' are linked to genes directly associated with oleate or linoleate synthesis because the alleles had opposite effects for oleate and linoleate content. Both 'B' G. max and 'A' G. soja alleles were associated with high linoleate and low oleate. The effect of these loci on other fatty acids probably was a secondary consequence of this primary effect. Loci on other linkage groups also were associated with each fatty acid. Although Graef et al. (1985) and Wilcox and Cavins (1985) reported mutants at individual loci that substantially altered the content of fatty acids in seed, many genes are involved in fatty acid synthesis (Slack and Browse, 1984). The significant markers may be linked to any of these genes.

Suarez (1989) mapped seed component traits in two G. max by G. soja populations. Three of the isozyme markers Suarez found significantly associated to the seed traits also were scored in the population 1. In this population, no associations were observed between these isozyme markers and the seed traits. An explanation for this lack of consistency between our findings and those of Suarez is that the marker-trait associations observed by Suarez may be population specific. This is not unexpected, for it has been observed for soluble solids in tomato (Lycopersicon esculentum) (Tanksley and Hewitt 1988) and most of the significant associations Suarez found were specific to only one of the populations he studied. Another explanation for the lack of consistency is that population 1 included fewer individuals than the populations used by Suarez. The smaller population resulted in lower sensitivity for finding significant marker-trait associations in our study than for Suarez. Therefore, the same associations found by Suarez may be present in our population, but the associations were too small to be declared significant.

One major goal of RFLP research is the improvement of cultivars through marker-assisted selection. G. soja is thought to be the progenitor of G. max and a repository of genetic diversity that could be utilized for soybean improvement (Hymowitz and Singh 1987). G. soja would be most useful in the improvement of G. max for traits that a large proportion of the variation is controlled by a few major genes that could be backcrossed into breeding material. Seed protein content may be such a trait. It was controlled in population 1 by loci in two different

linkage groups that explain 42 and 24% of the total variation for protein. Although great diversity exists within G. max for seed protein content (Burton, 1985), the G. soja alleles for high protein may not be currently available in G. max breeding material and could be incorporated into G. max to improve its protein content.

This study allowed us to find chromosomal regions that were significantly associated with seed component traits. These associations were tested in a second population and the findings were inconclusive. Further study will be required in this experiment to resolve the discrepancies between populations. Once this question is resolved other populations will be studied to find QTL in them. Information on markers in these and other populations should help breeders in constructing allelic combinations for further study and for the development of superior genotypes.

REFERENCES

Burton, J. W. 1985. Breeding soybeans for improved protein quantity and quality. p. 361-367. In R. Shibles (ed.) World soybean research conference III: Proceedings. Westview Press, Boulder CO.

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

Erickson, L. R., H. D. Voldeng, and W. D. Beversdorf. 1981. Early generation selection for protein in Glycine max x G. soja crosses. Can. J. Plant Sci. 61:901-908.

Ertl, D. S., and W. R. Fehr. 1985. Agronomic performance of soybean genotypes from Glycine max x Glycine soja crosses. Crop Sci. 25:589-592.

Fehr, W. R. 1987. Principles of Cultivar Development. Macmillan Publishing Company, New York.

Graef, G. L., W. R. Fehr, and S. R. Cianzio. 1989. Relation of isozyme genotypes to quantitative characters in soybean. Crop Sci. 29:683-688.

Graef, G. L., W. R. Fehr, and E. G. Hammond. 1985. Inheritance of three stearic acid mutants of soybean. Crop Sci. 25:1076-1079.

Howell, R. W., C. A. Brim, and R. W. Rinne. 1972. Plant geneticists' contribution toward changing lipid and amino acid composition of soybeans. J. Am. Oil Chem. Soc. 49:30-32.

Hymowitz, T, and R. J. Singh. 1987. Taxonomy and speciation. In J. R. Wilcox (ed.) Soybean: Improvement, production and uses. 2nd ed. Agronomy 16:23-48.

Keim, P, and R. C. Shoemaker. 1988. Construction of a random recombinant DNA library that is primarily single copy sequences. Soybean Genet. Newsl. 15:147-148.

Keim, P., R. C. Shoemaker, and R. G. Palmer. 1989. Restriction fragment length polymorphism diversity in soybean. Theor. Appl. Genet. 77:786-797.

Keim, P., B. W. Diers, and R. C. Shoemaker. 1990a. Genetic analysis of soybean hard seededness with molecular markers. Theor. Appl. Genet. 79:465-469.

Keim, P, B. W. Diers, T. Olson, and R. C. Shoemaker. 1990b. RFLP mapping in soybean: Association between marker loci and variation in quantitative traits. Genetics 126:735-742.

Lander, E. S., and D. Botstein. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185-199.

Lander, E. S., P. Green, J. Abrahamson, A. Barlow, M. J. Daly, S. E. Lincoln, and L. Newburg 1987. Mapmaker: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174-181.

Osborn, T. C., D. C. Alexander, and J. F. Fobes. 1987. Identification of restriction fragment length polymorphisms linked to genes controlling soluble solids content in tomato fruit. Theor. Appl. Genet. 73:350-356.

Palmer, R. G., and T. C. Kilen. 1987. Qualitative genetics and cytogenetics. In J. R. Wilcox (ed.) Soybean: Improvement, production and uses. 2nd ed. Agronomy 16:135-209.

Paterson, A. H., E. S. Lander, J. D. Hewitt, S. Peterson, S. E. Lincoln, and S. D. Tanksley. 1988. Resolution of quantitative traits into mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721-726.

Rennie, B. D., M. L. Thorpe and W. D. Beversdorf. 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.

Slack, C. R., and J. A. Browse. 1984. Synthesis of storage lipids in developing seeds. p. 209-244. In D. R. Murray (ed.) Seed Physiology Vol. 1 Development. Academic Press Australia, North Ryde, N.S.W.

Smith, K. J., and W. Huyser. 1987. World distribution and significance of soybean. In J. R. Wilcox (ed.) Soybeans: Improvement, production, and uses. 2nd ed. Agronomy 16:23-48.

Suarez, J. C. 1989. Association between isozyme genotypes with quantitative traits in soybean. Ph.D. Thesis. Iowa State University, Ames. University Microfilms International, Ann Arbor, Mich. (Diss. Abstr. 89-20190).

Tanksley, S. D., and J. Hewitt. 1988. Use of molecular markers in breeding for soluble solids content in tomato-a reexamination. Theor. Appl. Genet. 75:811-823.

Weber, C. R. 1950. Inheritance and interrelation of some agronomic and chemical characters in an interspecific cross in soybeans, Glycine max x G. ussuriensis. Iowa Agric. Exp. Stn. Res. Bull. 374.

Wilcox, J. R. 1985. Breeding soybeans for improved oil quantity and quality. p. 380-386. In R. Shibles (ed.) World soybean research conference III: Proceedings. Westview Press, Boulder CO.

Wilcox, J. R., and J. F. Cavins. 1985. Inheritance of low linolenic acid content of the seed oil of a mutant in Glycine max. Theor. Appl. Genet. 71:74-78.

Wilson, R. F. 1987. Seed metabolism. In J. R. Wilcox (ed.) Soybeans: Improvement, production, and uses. 2nd ed. Agronomy 16:643-686.

SECTION II: MAPPING OF PHYTOPHTHORA RESISTANCE LOCI IN SOYBEAN WITH RESTRICTION FRAGMENT LENGTH POLYMORPHISM MARKERS

ABSTRACT

Phytophthora root rot caused by Phytophthora megasperma Drechs. f. sp. glycinea Kuan and Erwin is one of the most serious diseases of soybean [Glycine max (L.) Merr.]. Six loci with alleles giving race-specific resistance of soybean to phytophthora have been reported. The first objective of this study was to map the phytophthora resistance loci using restriction fragment length polymorphism (RFLP) markers. The second objective was to map the Rj2 locus for ineffective nodulation with RFLP markers because of the linkage of Rj2 to Rps2. The mapping was conducted using a series of 'Williams' near-isogenic lines (NILs) each having one or two phytophthora resistance alleles. The NILs were screened with 141 mapped RFLP markers. At least one polymorphism was found between each NIL and the recurrent parent. Linkage tests among the polymorphic RFLP markers, Rps loci and the Rj2 locus were conducted using F_2 populations. Linkage was found between RFLP markers and Rps1, Rps2, Rps3, Rps4, Rps5 and Rj2.

Additional index words: Glycine max (L.) Merr., Phytophthora megasperma Drechs, f. sp. glycinea Kuan and Erwin, RFLP markers, Genetic mapping.

INTRODUCTION

Phytophthora root rot is one of the most serious diseases of soybean (Athow, 1987). There have been six loci reported with alleles that give race-specific resistance of soybean to phytophthora. Bernard et al. (1975) reported resistance at the Rps1 locus, Kilen et al. (1974) reported Rps2, Mueller et al. (1978) Rps3, Athow et al. (1980) Rps4, Buzzell and Anderson (1981) Rps5, and Athow and Laviolette (1982) Rps6. At some loci, more than one resistance allele have been reported (Athow, 1987).

Two Rps loci have been placed on the classical linkage map. Rps1 was linked to the metribuzin sensitivity locus (hm) and formed linkage group 10 (Kilen and Barrentine, 1983). Rps2 was linked to Rj2, a locus conferring effective versus ineffective nodulation to some strains of Bradyrhizobium japonicum and formed linkage group 19 (Devine et al., 1991).

Restriction fragment length polymorphism (RFLP) markers are a powerful tool in genetic mapping. The markers are cloned DNA segments that can be used to reveal base pair changes or rearrangements in homologous DNA sequences (Beckmann and Soller, 1983; Burr et al., 1983). The use of RFLP markers has led to the rapid construction of linkage maps in many plant species (Burr, 1989) and the placement of genes that control qualitative and quantitative traits onto these maps.

Genes controlling qualitative traits in plants can be mapped using near-isogenic lines (NILs) (Muehlbauer et al., 1989 and 1990; Young et al., 1988). A NIL can be formed by backcrossing a gene of interest from a donor parent into the genetic background of a recurrent parent. After several backcrosses, the genome of the NIL becomes nearly identical to the recurrent parent. According to formulas summarized by Muehlbauer et al. (1988), a soybean NIL formed through six backcrosses would have, on average, 2.1% of its genome originating from the donor parent. Of the contribution of the donor parent, 65% would be in a linkage block surrounding the gene backcrossed into the NIL. (The calculation assumes a genome of 20 chromosomes with each having a length of 100 centiMorgans [cM].) Therefore, markers that are polymorphic between the NIL and the recurrent parent have a high probability of being linked to the gene backcrossed into the NIL. Linkage between the polymorphic markers and the gene backcrossed into the NIL would be tested in populations segregating for both the marker and gene. NILs are available for many genes with qualitative effects in soybean including all six Rps loci (Bernard, 1976).

Several genes with qualitative effects have been mapped using NILs. Young et al. (1988) mapped two RFLP markers tightly to the Tm-2a locus in tomato (Lycopersicon esculentum) using NILs. Gardiner et al. (1989) used NILs to map the Ht1 locus in maize (Zea mays L.) with RFLP markers. Using NILs in soybean, the ln and dt_2 genes were mapped with isozymes (Muehlbauer et al. 1989) and the r and lf genes were mapped with RFLP markers (Muehlbauer et al. 1990).

Progress in classical genetic mapping in soybean has been slow. Palmer and Kilen (1987) reported a soybean linkage map that consisted of 40 markers and covered 420 cM. In contrast, the use of RFLP markers in soybean mapping has resulted in the rapid construction of maps with hundreds of markers (Keim et al., 1990; Tingey et al., 1989). Agronomically important genes should be included on RFLP maps to best utilize their potential. Because of the importance of phytophthora as a soybean disease, mapping Rps loci would be useful for both plant breeding and molecular studies. We report RFLP mapping of five Rps loci and the Rj2 locus in soybean. The loci were mapped by first identifying RFLP markers that detected polymorphisms between the NILs and the recurrent parent. These markers were then tested for linkage with the Rps loci and the Rj2 locus.

MATERIALS AND METHODS

Genetic material

The source of the six Rps resistance loci and the Rj2 ineffective nodulation locus used in this experiment was a series of 'Williams' NILs obtained from Dr. R. L. Bernard, the former curator of USDA Northern Soybean Germplasm Collection. Williams contained the allele conferring susceptibility at all six resistance loci and the recessive allele for effective nodulation. The genotypes and pedigrees of the NILs are summarized in Table 1.

The isolines and Williams were crossed to form F_2 populations for testing linkage among the RFLP markers, Rps loci and Rj2. The crosses were made in a greenhouse during the fall of 1988. The resulting F1 plants were grown during the spring of 1989 in a greenhouse. Seventy F_2 plants from each cross were then grown in a field near Ames, IA during the summer of 1989. The plants were allowed to naturally self pollinate and at maturity, each F_2 plant was threshed individually to form F_2-derived lines in the F_3 generation.

RFLP analysis

The RFLP genotypes were determined using DNA extracted from leaves. The leaf material for Williams and the isolines was obtained from seedlings grown in a greenhouse. The leaf material used to determine the genotypes of the F_2 individuals was obtained either from field-grown F_2 plants or from bulked leaf material harvested from six or more F_3 seedlings derived from an individual selfed F_2 plant. The F_3

seedlings were grown in a greenhouse. The field-grown material was harvested, frozen on dry ice in the field, then brought to the lab, dipped in liquid nitrogen and lyophilyzed. The greenhouse-grown material was harvested, frozen in liquid nitrogen and lyophilyzed.

DNA extractions were done according to the protocol for green-house-grown material described by Keim et al. (1988) with the addition of an extra purification procedure. After the DNA was precipitated and resuspended in Tris-EDTA buffer (TE) with pH 8.0, one-half volume of 7.5 M ammonium acetate was added. The solution was then mixed, incubated at room temperature for 30 minutes, and centrifuged for 20 minutes at 15,600xg. The supernatant was poured into new tubes and the precipitate was discarded. The DNA in solution was precipitated with 2.5 volumes of 95% ethanol, spooled with a glass rod into a new 1.5 ml microcentrifuge tube and washed with 80% ethanol. The DNA was dryed and resuspended in TE . The DNA was then quantified and digested with restriction enzymes.

The NILs and Williams were screened with 141 mapped recombinant clones to survey for DNA polymorphisms between the NILs and Williams. The clones were screened either in pairs (two recombinant clones were ^{32}P labelled together and hybridized to the same membrane) or singly. The number of enzymes used to cut the DNA of the NILs and Williams for the screening varied among clones. Some clones were hybridized to DNA cut with only the enzymes that gave polymorphisms for that clone between the G. max and G. soja parents used by Keim et al. (1990). Other clones were hybridized to

DNA from the NILs and Williams cut with each of the enzymes DraI, EcoRI, EcoRV, HinDIII, and TaqI. When an Rps locus was mapped with one marker, RFLP markers closely linked to the locus were screened for polymorphisms between the recurrent parent and the NIL using DNA cut with up to 20 enzymes. The hybridization protocol has been described previously (Apuya et al., 1988; Keim et al., 1989).

Phytophthora resistance testing

The genotype of the F_2 plants, for phytophthora resistance, was determined by progeny testing seven to 15 F_3 seedlings from each selfed F_2 plant. A modified hypocotyl-puncture method for inoculation (Morgan and Hartwig, 1965) was used for disease testing. Differential checks were included in each screening to test the purity of the phytophthora race used as inoculum.

The testing was done in a greenhouse by growing the seedlings in clay pots filled with sterile soil. When the unifoliolates began to expand, the seedlings were inoculated with a culture of phytophthora. The hypocotyls were slit 0.5 cm below the cotyledons and fungal mycelia were introduced into the slit. After inoculation, the seedlings were placed in a misting chamber for two days under reduced light conditions and then placed in the greenhouse. The seedlings were classified as either susceptible or resistant three to four days after inoculation. The diseased plants were withered and had necrosis spreading from the inoculation slit whereas the resistant plants were healthy and had scar tissue formed over the inoculation slit. The temperature in the greenhouse did not exceed 30°C in the day nor

below 21°C at night. Expected disease reactions were observed for the differential check cultivars for each test reported. Based on F_3 progeny data, the genotype of each F_2 plant was inferred. No attempt was made to distinguish between the homozygous resistant and heterozygous F_2 plants in calculating linkages. This was done because occasional resistant individuals were expected to give susceptible reactions making the homozygous and heterozygous classes very difficult to separate without using very large family size (40 or more individuals for each F_2).

Nodulation testing

The genotype of each F_2 plant for the Rj2 locus was determined by testing four to six F_3 plants derived from each F_2 plant for nodulation response. As with phytophthora resistance, no attempt was made to distinguish between the families derived from heterozygous and homozygous dominant F_2s. The seed were first surface sterilized with 5% Na hypochlorite (Chlorox®) for 10 minutes, quickly rinsed five times with sterile water (200 ml per rinse), and immediately sown in 23 or 25 cm diameter pots filled with sterile vermiculite. Prior to sowing, the pots were soaked overnight in a bleach solution, drained, lined with sterile germination paper, and filled with autoclaved vermiculite. The seeds were sown in the vermiculite and then each seed was inoculated with 1-ml of a culture containing 10^8 cells ml^{-1} of B. japonicum strain USDA 122. The culture was provided by Dr. Renee Kosslak and was grown in a Hepes-Mes buffered medium containing 1 g L^{-1} of arabinose, D-gluconic acid, and yeast extract (AIEHM [AG]

medium, [Kuykendall, 1979]). The pots were then covered with cellophane and placed in a darkened growth chamber. The chambers were activated soon after germination, and holes were cut in the cellophane to allow the seedlings to emerge from pots. The chambers were set on a 14-hour day length with a temperature of 28°C during the day and 22°C during the night. A saucer was place below each pot and the plants were watered as needed by filling the saucers with tap water.

At 12, 19, and 24 days post planting, the plants were fertilized with 150 ml of 5 X modified Hoagland solution which was free of fixed nitrogen (Imsande and Ralston, 1981). The only source of external N to the plants was from N_2 fixation. The plants were grown in two experiments with half of the families grown in each. Thirty-one days post planting the families were scored for nodulation response. The plants that were green were scored as effective nodulators (rj2/rj2 genotype) and the plants that were yellow were scored as ineffective nodulators (Rj2/rj2 or Rj2/Rj2) (Devine and O'Neill, 1989). The roots were checked for nodules to confirm the observations of the leaves. The green plants had effective nodules, which appeared normal, and the yellow plants had ineffective nodules, which appeared as collars at the base of branch roots.

Segregation analyses

Markers found to be polymorphic between NILs and Williams were tested for linkage with the resistance loci and the Rj2 nodulation locus in F_2 populations. The segregation of alleles at each locus was

tested to determine its fit to expected ratios. Linkage analyses between the loci were done using maximum likelihood with the program 'Linkage 1' (Suiter et al., 1983). When data were collected for three or more loci in a linkage group, the program Mapmaker (Lander et al., 1987) was used to determine the most probable order of loci using multipoint linkage analysis.

RESULTS

At least one marker revealed a polymorphism between each NIL and Williams (Table 1). To determine if genetic linkage existed between the Rps loci, the Rj2 locus and polymorphic marker loci, these were scored in several F_2 populations (Table 2). The Rps loci and Rj2 fit a 3:1 segregation ratio (Table 3). All markers that were scored as having dominant and recessive alleles did not differ significantly from a 3:1 segregation ratio and the markers that were scored as codominant did not differ significantly from a 1:2:1 ratio (Table 3). The linkage groups the loci map to are labelled according to Keim et al. (1990). Because markers are continually being added to the map, the labels for the linkage groups will likely change in the future.

Rps1

Four markers were polymorphic between Williams and L77-1794 (the NIL containing the resistance allele Rps1-k) and five between Williams and L81-4352 (the NIL containing the resistance alleles Rps1-c, Rps2 and Rj2) (Table 1). Segregation data were obtained for Rps1-k, and the markers pA-280, pK-418, and pK-395 in cross 1 (Table 3). All tested markers were significantly ($P<0.01$) linked to Rps1 except pK-395 (Table 4). The F_2s were scored for Rps1 and pK-395 without distinguishing between dominant homozygotes and heterozygotes (Table 3) and the dominant alleles for the two loci were in repulsion. In this situation, linkage tests are inefficient (Hanson, 1959). This is the likely reason for the nonsignificant linkage even though no individuals

Table 1. Near isogenic lines used in study, their pedigrees and the RFLP markers that were polymorphic between the NILs and Williams

Gene(s)	Line No.	Parentage	Polymorphic RFLP markers
Rps1-k	L77-1794†	Wm(7)‡ x Kingwa	pK-418, pK-395, pA-280, pA-71
Rps2, Rj2	L76-1988	Wm(6) x Harosoy(5) x D54-2437	pA-199, pA-233
Rps3	L83-569	Wm(6) x PI 86.972-1	pA-186, pA-757, pR-45
Rps4	L85-2350	Wm(6) x PI 86.050	pT-5, pA-586
Rps5	L85-3044	Wm(6) x PI91.160	pT-5
Rps6	L88g-52	Wm(5) x Altona	pK-258
Rps1-c, Rps2, Rj2	L81-4352	Wm79 x L76-2013	pK-418, pK-395, pA-280, pA-71, pA-199, pA-233

†L77-1794 is 'Williams82.'

‡Number of backcrosses used in developing the NILs.

Table 2. Crosses used to test linkages

Cross No.	Parents	Genes
1	L77-1794 x L85-3044	Rps1-k, Rps5
2	L81-4352 x Williams	Rps1-c, Rps2, Rj2
3	L83-569 x Williams	Rps3
4	L85-2350 x Williams	Rps4
5	L88g-52 x L85-3044	Rps6, Rps5

Table 3. Segregation ratio of resistance genes and markers

Locus	Race or Enzyme†	Cross No.	Expected Ratio	Observed Ratio	n	χ^2	P‡
Rps1	7	1	3:1	51:9	60	3.20	0.07
pK-395	HinDIII	1	3:1	33:11	44	0.00	1.00
pA-280	DraI	1	1:2:1	17:32:9	58	2.83	0.24
pK-418	PstI, EcoRI	1	1:2:1	19:32:7	58	5.59	0.06
Rps2	4	2	3:1	45:18	63	0.43	0.52
Rj2	USDA 122	2	3:1	40:12	52	0.01	0.75
pA-199B	TaqI	2	1:2:1	14:30:15	59	0.05	0.97
pA-233	BclI	2	1:2:1	11:24:16	51	1.16	0.56
Rps3	4 or 25	3	3:1	42:20	62	1.74	0.19
pA-186	DraI	3	1:2:1	12:26:18	56	1.57	0.46
pR-45	BclI	3	1:2:1	11:29:20	60	2.77	0.25
Rps4	4 or 25	4	3:1	32:12	44	0.12	0.73
pT-5	HinDIII	4	3:1	29:11	40	0.13	0.72
pA-586	EcoRI	4	1:2:1	11:19:13	43	0.77	0.68
Rps5	25	1	3:1	52:12	64	1.33	0.25
pT-5	HinDIII	1	1:2:1	15:30:11	56	0.86	0.65

†Phytophthora race used to test for resistance, strain of rhizobium used to test nodulation or enzyme used to cut the genomic DNA to score marker.

‡Probability of a greater value of χ^2.

Table 4. Linkage data for polymorphic markers and Rps resistance loci and Rj2 locus

Cross No.	Genotypes	Genotypic Classes†									no.	χ^2	Pr‡	P§	SE¶
		e	f	g	hi	j	k	l	m	n					
1	pA-280/pK-418	16	2	1	28	0	0	0	1	7	55	87.79	0.000	3.7	1.8
2	pA-199/pA-233	9	2	1	20	0	2	0	0	14	48	68.90	0.000	5.4	2.4
3	pA-186/pR-45	9	2	3	22	0	2	0	0	16	56	64.50	0.000	8.4	2.7
		ef	ghi	jk	l	m	n								
1	Rps1/pA-280	17	30	1	0	1	7				56	40.95	0.000	4.3	2.8
1	Rps1/pK-418	19	30	0	0	1	7				57	48.98	0.000	2.2	2.0
2	Rps2/pA-199	14	27	0	0	3	15				59	46.26	0.000	4.9	2.9
2	Rps2/pA-233	11	23	0	0	1	16				51	46.69	0.000	1.8	1.9
2	Rj2/pA-233	10	22	0	0	0	12				44	44.00	0.000	0.0	ND#
2	Rj2/pA-199	12	24	0	0	2	10				48	38.15	0.000	4.3	3.0
3	Rps3/pA-186	12	24	2	0	2	16				56	39.38	0.000	6.5	3.4
3	Rps3/pR-45	11	28	1	0	1	19				60	51.38	0.000	2.9	2.2
4	Rps4/pA-586	11	19	1	0	0	12				43	38.41	0.000	2.3	2.3
4	pT-5/pA-586	10	16	3	0	1	10				40	23.70	0.000	9.8	4.9
1	Rps5/pT-5	15	30	0	0	0	11				56	56.00	0.000	0.0	ND

		a	b	c	d					
2	Rps2/Rj2	40	0	0	12	52	52.00	0.000	0.0	ND
4	Rps4/pT-5	26	2	3	9	40	19.40	0.000	12.6	5.7
		a'	b'	c'	d'					
1	Rps1/pK-395	26	11	7	0	44	2.77	0.096	0.0	ND

†Class designations per Allard (1956). a=A_B_, b=A_bb, c= aaB_, d=aabb, a'=A_b_, b'=A_BB, c'=aab_, d'=aaBB, e=AABB, f=AaBB, g=AABb, hi=AaBb, j=AAbb, k=Aabb, l=aaBB, m=aaBb, n=aabb. A=allele from the resistant parent, a=allele from the susceptible parent from the first locus listed, and B= allele from the resistant parent and b=allele from the susceptible parent for the second locus listed.

‡Probability of a greater value of χ^2.

§Estimate of frequency of recombination.

¶Standard error of recombination estimate.

#Standard error of recombination estimate not defined.

in the recombinant class were observed. Because of this, pK-395 was not considered further.

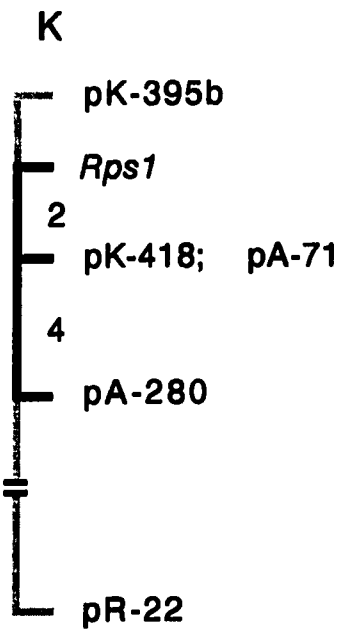
The order of the markers and Rps1 was determined using the program Mapmaker (Lander et al., 1987). The most probable order was shown to be pA-280, pK-418 and Rps1 (Figure 1). The data from cross 1 combined with mapping data from the G. soja by G. max population (Keim et al., 1990) indicated that Rps1 was between pK-395 and pK-418 (Figure 1). Rps1 is likely on linkage group K (Keim et al., 1990) because pA-280 and pK-418 are on this group.

Rps2

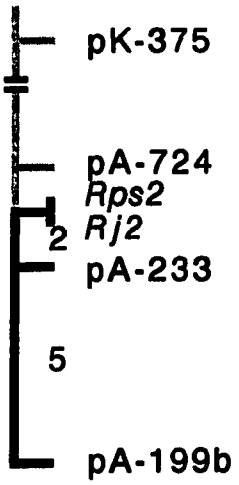
Polymorphisms with markers pA-199, and pA-233 were found between Williams and L76-1988 (the NIL with Rps2 and Rj2) and L81-4352 (the NIL with Rps1, Rps2 and Rj2) (Table 1). Segregation data were obtained for Rps2, Rj2, pA-199 and pA-233 in cross 2 (Table 3). Linkage analysis showed that all four loci were significantly ($P<0.01$) linked (Table 4). These results indicate that Rps2 and Rj2 are on linkage group L because this group contains pA-233 and pA-199b (Keim et al., 1990).

The most probable order of the loci, based on Mapmaker analysis, was pA-199b, pA-233 and Rps2 with the position of Rj2 ambiguous (Figure 1). Three positions of Rj2 have essentially the same likelihood. Rj2 could either be between pA-199 and pA-233, pA-233 and Rps2 or Rps2 and pA-724.

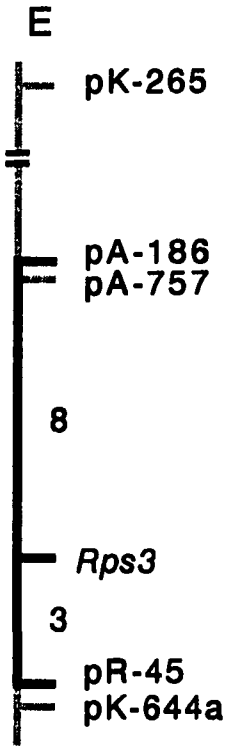
Rps1 Linkage



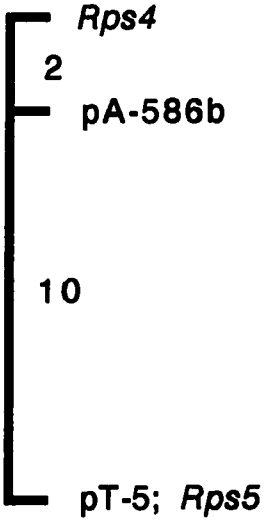
Rps2 Linkage



Rps3 Linkage



Rps4 and *Rps5* Linkage



Rps3

Polymorphisms with markers pA-186, pA-757 and pR-45 were found between Williams and L83-569 (the NIL with Rps3) (Table 1). Segregation analysis was done for pA-186, pR-45 and Rps3 in population 3 (Table 3) All three were found to be significantly linked ($P<0.01$) (Table 4). pA-186 was on linkage group E (Keim et al., 1990) and has been linked with pR-45 and pA-757 in the G. max by G. soja population (unpublished data). This indicates that Rps3 also was on group E. The order of the markers and Rps3 was obtained using the program Mapmaker. The most probable order is pA-186, Rps3, and pR-45 (Figure 1).

Rps4 and Rps5

Polymorphisms with the markers pT-5, and pA-586 were found between Williams and L85-2350 (the NIL with Rps4) (Table 1). Linkage analysis showed both pT-5 and pA-586 were linked to Rps4 (Tables 3 and 4). Mapmaker was used to order the loci. The most likely order is Rps4, pA-586 and pT-5 (Figure 1).

Neither pT-5 or pA-586 were mapped by Keim et al. (1990). Both have been mapped recently in the G. max by G. soja population (unpublished data). The marker pA-586 detected two RFLP loci that were unlinked in this population. A molecular weight comparison of the polymorphic fragment that was linked to Rps4 with the fragments that mapped each allele of pA-586 in the G. max by G. soja population indicated that the polymorphic fragment linked to Rps4 was pA-586b.

A polymorphism with marker pT-5 was found between Williams and L85-3044, the (NIL with Rps5) (Table 1). No recombinants were observed between pT-5 and Rps5 (Table 4).

These data suggested that Rps4 and Rps5 also were linked (Figure 1). Although a direct linkage test between Rps4 and Rps5 was not done because a population segregating for the two loci was not available, both were linked to the same locus of pT-5 (Table 4, Figure 1). The markers pA-586b and pT-5 were not linked in the G. max by G. soja population (unpublished data). However, aberrant segregation was obtained for pT-5 in that population which made mapping of that marker difficult.

Rps6

A polymorphism with marker pK-258 was observed between Williams and L88g-52 (the NIL for Rps6). The marker was tested against DNA isolated from population 5, which was determined by phytophthora screening to be segregating for Rps6 (Table 2). Segregation for pK-258 was not observed in this population. A possible explanation as to why pK-258 was not segregating was that the seed we received for L88g-52 was from a BC5F1 plant and therefore segregating for resistance. The gamete from L88g-52 that was used to form cross 5 (Table 2) may have contained the Rps6 allele but not the donor parent allele for pK-258. Although a linkage test could not be completed between Rps6 and pK-258, these results are an indication that the two loci are unlinked or at least loosely linked. If tight linkage between the two existed, it would be unlikely for the Rps6 resistance allele and pK-

258 allele from L88g-52 to be separated by recombination and not be in the same gamete.

DISCUSSION

Five Rps loci conferring resistance to phytophthora in soybean and the Rj2 nodulation locus have been mapped with RFLP markers. The most probable linkage orientation of the loci are given on Figure 1. Because of the tight linkages found and the relatively small families used in this study, there were few individuals with recombination between the markers and the genes. Therefore errors in determining the linkage orientations among the loci may have occurred so the orientations should be considered tentative. A major concern at the start of the project was that the level of DNA polymorphism between the recurrent parent and the NILs would be insufficient for successful mapping. However, the level was sufficient, which is consistent with the results of Muehlbauer et al. (1989, 1990).

Linkages were found between the Rps loci and all polymorphic markers from which F_2 data were obtained except for pK-258 that was polymorphic for the Rps6 NIL. The probability of linkage between a marker that is polymorphic between the donor and recurrent parent and a gene backcrossed into a NIL is 0.65, assuming six backcrosses, 20 chromosomes and 2,000 cM (Muehlbauer et al. 1988). We could conclude that linkages were found five times out of six because when more than one marker was polymorphic for a Rps locus, these markers identified the same linkage block and could not be considered independent. This frequency is higher than expected from the theory, but not significantly. An accurate estimate of the distribution and

amount of donor parent genome present in the NILs was not obtained. This was because regions of donor parent origin could have gone undetected since no markers that mapped to these regions were screened. Alternatively, markers mapping to these regions were monomorphic between the NIL and the recurrent parent.

Most soybean RFLP markers hybridize to several fragments of differing sizes. Keim et al. (1990) found that approximately 20% of the markers in the G. max by G. soja population were polymorphic for more than one fragment and that each polymorphic fragment mapped to a different locus. Therefore, each fragment could be assumed to be a different locus. Because of this, it is important to determine which locus is being mapped when comparing linkages to RFLP markers in different populations. This determination can be accomplished two ways. One way is to obtain segregation data for the locus being mapped with two markers that are known to have linked loci. This was done for mapping Rps1. Linkage was obtained between Rps1 and the markers pA-280 and pK-418. Both markers had loci that mapped to linkage group K, therefore, it is highly probable that Rps1 is also on linkage group K because the likelihood of mapping Rps1 to a second locus on a different linkage group for both markers is low. In the second method, the molecular weights of the polymorphic fragments used to map a marker were compared between populations. For each marker linked to an Rps locus, a comparison was made between the molecular weights of the polymorphic fragments used for mapping the Rps locus with the fragments used for mapping the marker in the G. max by G. soja

population. If the same fragments were polymorphic in both populations, this indicated that the same locus was mapped in both.

The linkage between Rps4 and Rps5 could be useful in backcrossing phytophthora genes into cultivars or breeding lines. The two genes in coupling could be backcrossed into a recurrent parent much more efficiently than using separate backcrossing programs for two unlinked genes. Using resistance alleles at the Rps4 and Rps5 loci together would provide a unique range of resistance that is unavailable for only one locus (Athow, 1987).

Classical linkage group 19 containing Rps2 and Rj2 has been mapped with RFLP markers. Our population segregating for Rps2 and Rj2 was too small to unambiguously orientate the genes with pA-233. The tight linkage found in this study between Rps2 and Rj2 agrees with the estimate of 2.7±1 recombination units between the loci obtained by Devine et al. (1991).

A potential use of RFLP markers in plant breeding is the selection of individuals based on the genotype of RFLP markers that are linked to important genes. For this to be practical, marker-assisted selection should be less expensive or tedious than direct selection of the trait. Testing of individuals for phytophthora resistance in soybean is relatively inexpensive and accurate. Therefore, marker-assisted selection is not practical for the plant breeder using standard RFLP technologies. However, new technologies are being developed that may allow fast and inexpensive DNA-based molecular screening. We are exploring the use of allele-specific oligonucleotides in conjunction with

polymerase chain reaction (PCR) as a method for fast molecular screening (Saiki et al., 1986; Oste, 1988). This technology includes the amplification of the DNA sequence of alleles using PCR reaction and then the identification of the allele using gel electrophoresis. This method could give a quick and inexpensive screening procedure that would be more efficient than the standard disease screen.

REFERENCES

Allard, R. W. 1956. Formulas and tables to facilitate the calculation of recombination values in heredity. Hilgardia 24:235-278.

Apuya, N., B. L. Frazier, P. Keim, E. J. Roth and K. G. Lark. 1988. Restriction length polymorphisms as genetic markers in soybean, Glycine max (L.) Merr. Theor. Appl. Genet. 75:889-901.

Athow, K. L. 1987. Fungal Diseases. In J. R. Wilcox (ed.) Soybean: Improvement production and uses. 2nd ed. Agronomy 16:687-727.

Athow, K. L., and F. A. Laviolette. 1982. Rps6, a major gene for resistance to Phytophthora megasperma f. sp. glycinea in soybean. Phytopathology 72:1564-1567.

Athow, K. L., F. A. Laviolette, E. H. Mueller, and J. R. Wilcox. 1980. A new major gene for resistance to Phytophthora megasperma var. sojae in soybean. Phytopathology 70:977-980.

Beckmann, J. S., and M. Soller. 1983. Restriction fragment length polymorphisms in genetic improvement: Methodologies, mapping and costs. Theor. Appl. Genet. 67:35-43.

Bernard, R. L. 1976. United States national germplasm collections. p. 286-289. In L. D. Hill (ed.) World Soybean Research. Interstate Printers and Publ. Inc., Danville, IL.

Bernard, R. L, P. E. Smith, M. J. Kaufmann, and A. F. Schmitthenner. 1975. Inheritance of resistance to phytophthora root and stem rot in the soybean. Agron J. 49:391.

Burr, B. 1989. Introduction. pp. 1-3. In T. Helentjaris and B. Burr (ed.) Development and application of molecular markers to problems in plant genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Burr, B., S. V. Evola, F. A. Burr, J. S. Beckmann. 1983. The application of restriction fragment length polymorphism to plant breeding. pp. 45-59. In J. K. Setlow and A. Hollaender (eds.) Genetic engineering principles and methods. Vol. 5. Plenum Press, New York.

Buzzell, R. I, and T. R. Anderson. 1981. Another major gene for resistance to Phytophthora megasperma var. sojae in soybean. Soybean Genet. Newsl. 8:30-33.

Devine, T. E. and J. J. O'Neill. 1989. Genetic allelism of nodulation response genes Rj1, Rj2 and Rj4 in soybean. Crop Sci. 29:1347-1350.

Devine, T. E., T. C. Kilen, and J. J. O'Neill. 1991. Genetic linkage of the phytophthora resistance gene Rps2 and the nodulation response gene Rj2 in soybean. Crop Sci. In Press.

Gardiner, J. M., D. A. Hoisington, and S. Melia-Hancock. 1989. Use of isogenic lines for mapping Ht loci in Zea mays L. p. 81. In Agronomy abstracts. ASA, Madison, WI.

Hanson, W. D. 1959. Minimum family sizes for the planning of genetic experiments. Agron. J. 51:711-715.

Imsande, J. and E. J. Ralston. 1981. Hydroponic growth and the nondestructive assay for dinitrogen fixation. Plant Physiol. 68:1380-1384.

Keim, P., T. C. Olson and R. C. Shoemaker. 1988. A rapid protocol for isolating soybean DNA. Soybean Genet. Newsl. 15:150-152.

Keim, P., R. C. Shoemaker and R. G. Palmer. 1989. Restriction fragment length polymorphism diversity in soybean. Theor. Appl. Genet. 77:786-792.

Keim, P., B. W. Diers, T. C. Olson, and R. C. Shoemaker. 1990. RFLP mapping in soybean: Association between maker loci and variation in quantitative traits. Genetics 126:735-742.

Kilen, T. C., and W. L. Barrentine. 1983. Linkage relationships in soybean between genes controlling reactions to phytophthora rot and metribuzin. Crop Sci. 23:894-896.

Kilen, T. C., E. E. Hartwig, and B. L. Keeling. 1974. Inheritance of a second major gene for resistance to phytophthora rot in soybeans. Crop Sci. 14:260-262.

Kuykendall, L. D. 1979. Transfer of R factors to and between genetically marked sublines of Rhizobium japonicum. Appl. Environ. Microbiol. 37:862-866.

Lander, E. S., P. Green, J. Abrahamson, A. Barlow, M. J. Daly, S. E. Lincoln, and L. Newburg. 1987. Mapmaker: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174-181.

Morgan, F. L., and E. E. Hartwig. 1965. Physiologic specialization in Phytophthora megasperma var. sojae. Phytopathology 55:1277-1279.

Muehlbauer, G. J., J. E. Specht, M. A. Thomas-Compton, P. E. Staswick, and R. L. Bernard. 1988. Near-isogenic lines-A potential resource in the integration of conventional and molecular marker linkage maps. Crop Sci. 28:729-735.

Muehlbauer, G. J., J. E. Specht, P. E. Staswick, G. L. Graef, and M. A. Thomas-Comptom. 1989. Application of the near-isogenic line gene mapping technique to isozyme markers. Crop Sci. 29:1548-1553.

Muehlbauer, G. J., P. E. Staswick, J. E. Specht, G. L. Graef, R. C. Shoemaker, and P. Keim. 1990. RFLP mapping in the soybean (Glycine max L. Merr.) using near-isogenic lines. Theor. Appl. Genet. (In review).

Mueller, E. H., K. L. Athow, and F. A. Laviolette. 1978. Inheritance of resistance to four physiologic races of Phytophthora megasperma var. sojae. Phytopathology 68:1318-1322.

Oste, C. 1988. Polymerase chain reaction. Biotechniques 6:162-167.

Palmer, R. G. and T. C. Kilen. 1987. Qualitative genetics and cytogenetics. In J. R. Wilcox (ed.) Soybean: Improvement, production and uses. 2nd ed. Agronomy 16:135-209.

Saiki, R. K., T. L. Bugawan, G. T. Horn, K. B. Mullis and H. A. Erlich. 1986. Analysis of enzymatically amplified ß globin and HLA-DQa DNA with allele-specific oligonucleotide probes. Nature 324:163-166.

Suiter, K. A., J. F. Wendel and J. S. Case. 1983. Linkage-1: A pascal computer program for the detection and analysis of genetic linkage. J. Hered. 74:203-204.

Tingey, S., S. Sebastian, and J. A. Rafalski. 1989. A RFLP map of the soybean genome. p. 180. In Agronomy abstracts. ASA, Madison, WI.

Young, N. D., D. Zamir, M. W. Ganal and S. D. Tanksley. 1988. Use of isogenic lines and simultaneous probing to identify DNA markers tightly linked to the Tm-2a gene in tomato. Genetics 120:579-585.

SECTION III: GENETIC ANALYSIS OF A VARIANT OF THE A4 POLYPEPTIDE OF THE GY4 GLYCININ GENE IN SOYBEAN

ABSTRACT

The storage protein of soybean [Glycine max (L.) Merr.] seed is composed primarily of the globulin proteins glycinin and ß-conglycinin. Five subunits of glycinin have been described. Each subunit is composed of an acidic and basic polypeptide except for the subunit G4, which is composed of two acidic and one basic polypeptide. A previously unreported variant of the A4 polypeptide from the G4 subunit is described in this study. This variant was identified in PI 468916, the G. soja parent of the population being used to form a public RFLP map. The variant was observed using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) of seed storage protein. The A4 variant has a lower mobility than the A4 found in the G. max parent, A81-356022, or in any other genotype reported in the literature. The two forms of A4 segregated codominatly in the mapping population. The Gy4 gene that encodes A4 was mapped based upon the segregation in the F_2 mapping population of the variant polypeptide and upon a RFLP revealed through hybridization of a cDNA of Gy4 to restriction digested DNA of the F_2 individuals. The segregation of the variant polypeptide and the RFLP were identical. This confirms that the variant polypeptide is A4. It also suggests that the mobility difference was result of a change in the sequence of the gene that encodes A4 and not in an unlinked gene affecting the modification of A4. A germplasm survey of 24 G soja plant introductions also was conducted which revealed two genotypes with a mobility of A4 similar to PI 468916 and

two with mobilities faster than any other genotype reported in the literature.

INTRODUCTION

The storage protein in soybean [Glycine max (L.) Merr] seed is composed primarily of the globulin proteins glycinin and ß-conglycinin (Wilson, 1987). Because of the importance of these proteins, their structures have been studied extensively. Glycinin is an oligomer of six subunits (Badley et al., 1975). Five of the subunits have been identified based on their primary structure (Moreira et al., 1979). Each subunit is composed of an acidic and a basic polypeptide component linked by a disulfide bond, with exception of the subunit G4 (Staswick et al., 1984). G4 is composed of one basic and two acidic polypeptides.

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) is used to separate and visualize the polypeptides that compose the storage protein. Many of the polypeptides are not resolved with standard SDS-PAGE techniques, for the acidic glycinins $A1_a$, $A1_b$, $A2$ and $A4$ and the basic glycinins $B1_a$, $B1_b$, $B2$ and $B4$ usually comigrate as a group (Moreira et al., 1979). Fontes et al. (1984) included 5 M urea in the polyacrylamide gels and was able to separate $A4$ from the group of acidic glycinins.

Five glycinin genes have been cloned and studied. These genes are Gy1 to Gy5 and they encode glycinin subunits G1 to G5 respectively (Nielsen et al., 1989). The acidic and basic polypeptide for each subunit are transcribed together from the same gene and are post-translationally cleaved (Tumer et al., 1982). These genes have been classified into two groups based on their sequence homology and size.

The group 1 genes were identified by Fischer and Goldberg (1982) and include Gy1, Gy2 and Gy3. Scallon et al. (1985) identified the group 2 genes Gy4 and Gy5.

Mutants for glycinin storage protein genes have been described. The cultivar 'Raiden' was found to lack the G4 glycinin subunit encoded by the Gy4 gene (Staswick and Nielsen, 1983; Kitamura et al., 1980). Molecular studies later showed that the absence of the G4 subunit in Raiden was the result of a mutation in the initiation codon for the Gy4 gene (Scallon et al., 1987). The cultivar 'Forest' was found to lack the G3 subunit of glycinin. This null was caused by a chromosomal rearrangement which resulted in the separation of the two ends of the Gy3 gene that encodes the G3 subunit (Cho et al., 1989).

Staswick et al. (1983) conducted a germplasm survey to identify storage protein variants. They screened 120 lines of Glycine soja, an annual that shares the subgenus Soja with G. max, by comparing the mobilities of the storage protein components using SDS-PAGE. No obvious variants were found among the lines. Staswick et al. (1983) widened their search to include four perennial species of the subgenus Glycine. The species were G. canescens, G. tomentella G. tabacina and G. clandestina. A unique electrophoretic pattern for storage proteins was observed for each species. There were differences both in the number of bands and in their mobilities.

Efforts have been made to map the glycinin genes. Gy1 and Gy2 were shown to be tightly linked (Nielsen et al., 1989). Cho et al. (1989) studied the linkage of the other glycinin genes and found all to be

segregating independently. Five morphological markers also were tested and found unlinked to the glycinin genes.

RFLP markers allow efficient genetic mapping because more markers can be scored in a single population than with traditional methods. A soybean RFLP map was reported by Keim et al. (1990) that included 150 markers. Presently, this map includes an excess of 300 markers (Shoemaker, unpublished data). This map was constructed using a F_2 population formed from an interspecific cross between a G. max experimental line and a G. soja plant introduction.

An effort is being made to include markers of known biochemical functions on the public RFLP map. We have discovered a previously undescribed variant of the A_4 polypeptide component of the G4 glycinin subunit in the G soja parent of the mapping population. In this study, we describe the mapping and genetic analysis of the variant polypeptide.

MATERIALS AND METHODS

Genetic Material

The Gy4 gene was mapped using a F_2 population derived from a cross between the G. max experimental line A81-356022 and the G. soja Plant Introduction 468916. This population is being used to construct a public RFLP map. The F_2 seed for this population were obtained from Dr. W. R. Fehr, Iowa State University. Sixty F_2 plants were grown from this seed during the summer of 1987 near Ames, IA. Each F_2 plant was harvested and threshed separately to form F_2-derived lines in the F_3 generation ($F_{2:3}$). The $F_{2:3}$ lines were grown in a replicated trial during 1988 at three locations near Ames, IA (Keim et al., 1990). The $F_{2:4}$ seed from each plot were harvested in bulk. The F_2-derived families were scored twice for the protein variant using SDS-PAGE. The protein used was isolated from a bulk of four $F_{2:3}$ seed and a bulk of four $F_{2:4}$ seeds derived from each F_2 individual. DNA for RFLP analysis was isolated from leaf material harvested from the F_2 plants.

Protein Extraction

Protein was extracted by crushing seed between two sheets of glassine paper with a pestle. The crushed seed were then placed in a 1.5-ml microcentrifuge tube and soluble protein was extracted in 0.5-1.0 ml of extraction buffer. The protein used in gels without urea were extracted in a buffer containing 50 mM Tris (pH 8.0), 20 gL^{-1} SDS, and 10 mlL^{-1} ß-mercaptoethanol. The protein used in the gels containing urea were extracted in the same buffer with the addition of 5 M urea.

The samples were extracted for 15 min on ice with periodic mixing and then centrifuged for 10 min at 4°C at 15,000 x g. The supernatant was then pipetted into fresh tubes and quantified using the Biorad protein assay kit (Bio-rad Laboratories, Richmond, CA) with bovine serum albumin as a standard.

Protein Separation

A SDS-PAGE system adapted from Laemmli (1970) was used for protein separation. The resolving gels were 11 cm long and the stacking gels were 2 cm long. The electrophoretic buffer contained 0.025 M Tris, 1 g L^{-1} SDS and 0.1875 M glycine. Protein staining and destaining procedures were described by Honeycutt et al. (1989). The F_2-derived families were scored as having a F_2 genotype of homozygous G. max, heterozygous or homozygous G. soja.

Gels without urea The resolving gels were composed of a gradient of 80 to 160 gL^{-1} acrylamide with 0.44 M Tris (pH 8.8) and 1.2 gL^{-1} SDS. The stacking gels contained 55 gL^{-1} acrylamide with 0.125 M Tris (pH 6.8) and 1.0 gL^{-1} SDS. The polymerization of the resolving gels was catalyzed with 0.417 gL^{-1} ammonium persulfate and 0.167 mlL^{-1} TEMED (N, N, N', N'-tetramethylethylenediamine). The polymerization of the stacking gels was catalyzed with 0.625 gL^{-1} ammonium persulfate and 0.33 mlL^{-1} TEMED. The gels were run approximately 11 h at a constant 35 milliamps gel^{-1}.

Gels with urea. The resolving gels were composed of a gradient of 100 to 185 gL^{-1} acrylamide with 0.58 gL^{-1} SDS, 0.44 M Tris and 5 M urea. The stacking gels contained 50 gL^{-1} acrylamide, 0.44 M Tris, 0.58

gL^{-1} SDS and 5 M urea. Both the resolving and stacking gels were polymerized with 0.31 gL^{-1} ammonium persulfate and 0.167 mlL^{-1} TEMED.

RFLP Methods

The Gy4 gene was mapped using pG-248, a Gy4 cDNA clone (Dickinson et al., 1987). The clone was cut from the plasmid, gel purified and radiolabelled using random hexamer primers. The clone was then hybridized to HindIII-cut DNA of F_2 individuals from the mapping population. To determine the relative location of the HindIII polymorphism, the clone was cut from the plasmid and cut at an internal HindIII site. Each end was isolated by gel purification and used separately as a hybridization probe. The hybridization protocol was described previously (Apuya et al., 1988; Keim et al., 1989).

Mapping Gy4

F_2 segregation data were obtained both for pG-248 using a HindIII RFLP and the protein variant using SDS-PAGE without urea. These data were analyzed using the program 'Mapmaker' (Lander et al., 1987) together with an existing set of F_2 data consisting of over 300 makers.

Germplasm screen

Twenty-four G. soja plant introductions were screened with SDS-PAGE gels to determine if the variant polypeptide found in PI 468916 could be identified in these G. soja accessions. Protein was extracted from four seed of each introduction. The SDS-PAGE system used was the same as described for the F_2 population and parents.

RESULTS AND DISCUSSIONS

The parents of the population used for construction of a public RFLP map were screened for storage protein variation using SDS-PAGE without urea (Figure 1). A major polypeptide was found to be variant between the parents. A comparison of the parents with other cultivars with published storage protein mobilities (Fontes et al., 1984) revealed that the G. soja parent PI 468916 had a variant polypeptide that has not been described. This polypeptide had a lower mobility than found in other genotypes including the G. max parent A81-356022.

The variant polypeptide was identified as A4. This was determined because the variant corresponded to a missing polypeptide in the cultivar Raiden using SDS-PAGE systems both with and without urea (Figures 1 and 2). Although Raiden lacks the polypeptides A4, A5 and B3, it is clear that the variant polypeptide in PI 468916 is A4 because this is the only polypeptide lacking in Raiden that has a molecular weight similar to the variant. Generally A4 can not be resolved from A1a, A1b and A2 using SDS-PAGE without urea included in the gel (Fontes et al., 1984). However, by running the gels further than normal, we were able to separate A4 from the other acid polypeptides. This variant found in PI 468916 was the first mobility variant of A4 described.

Gy4, the gene that encodes A4, A5, and B3, was mapped using pG-248, a cDNA of Gy4 (Dickinson et al., 1987). The clone was first hybridized to parental DNA cut with the restriction enzymes BamH1,

Figure 1. Analysis of the storage proteins of parents and F_2-derived lines using SDS-PAGE. The electrophoretic conditions are given in the materials and methods section. The α', α and β are polypeptides of β-conglycinin. The A_3, A_4 and A's are acidic polypeptides of glycinins. The A's band consists of A_{1a}, A_{1b} and A_2. The genotypes of seed used in each lane are 1, Raiden; 2, A81-356022; 3, PI 468916; and 4, 5 and 6 are each from four seeds of F_2-derived families that are heterozygous, homozygous G. soja and homozygous G. max respectively

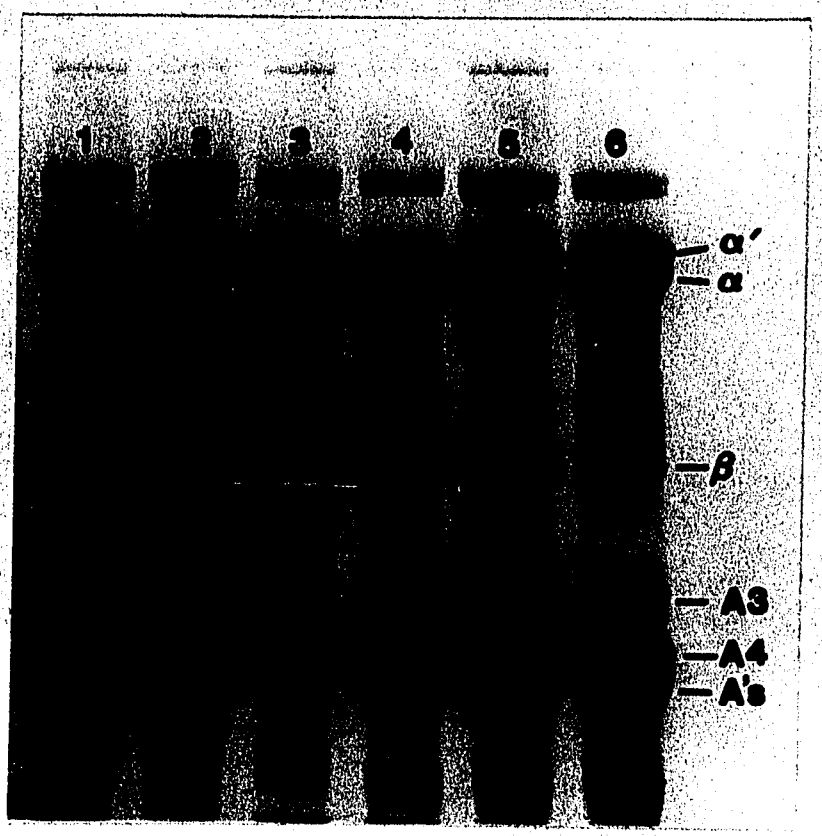
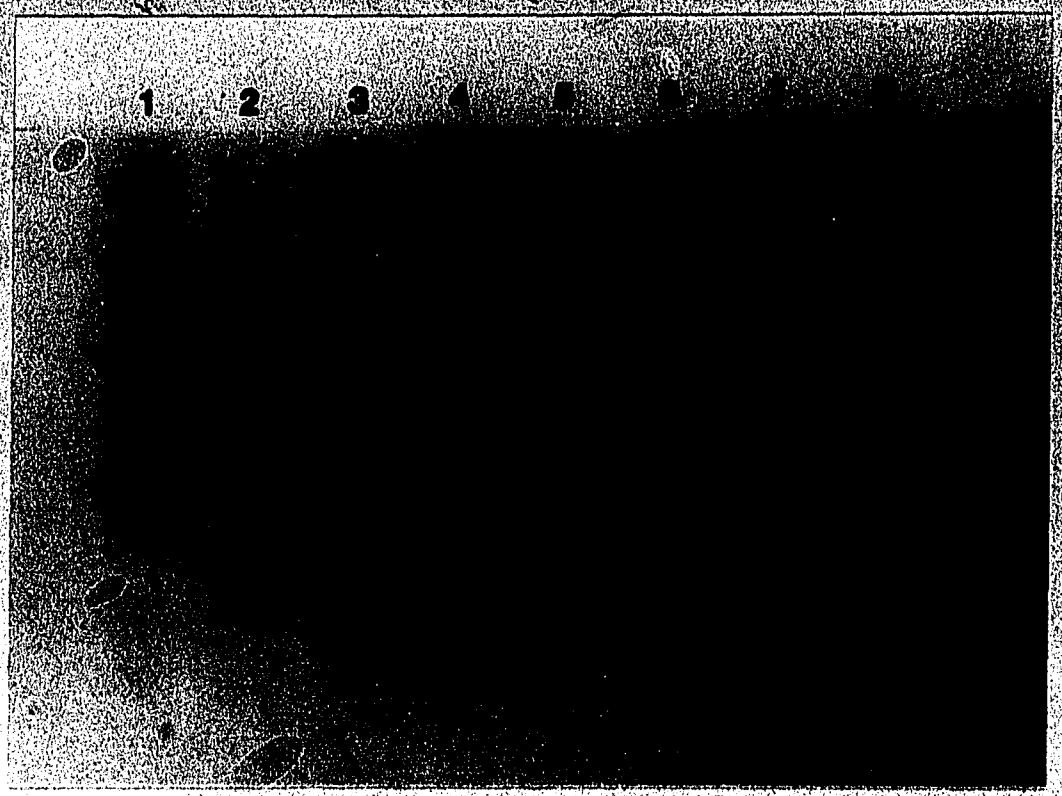

1
2
3
4
5
6
α′
α
β
A3
A4
A's

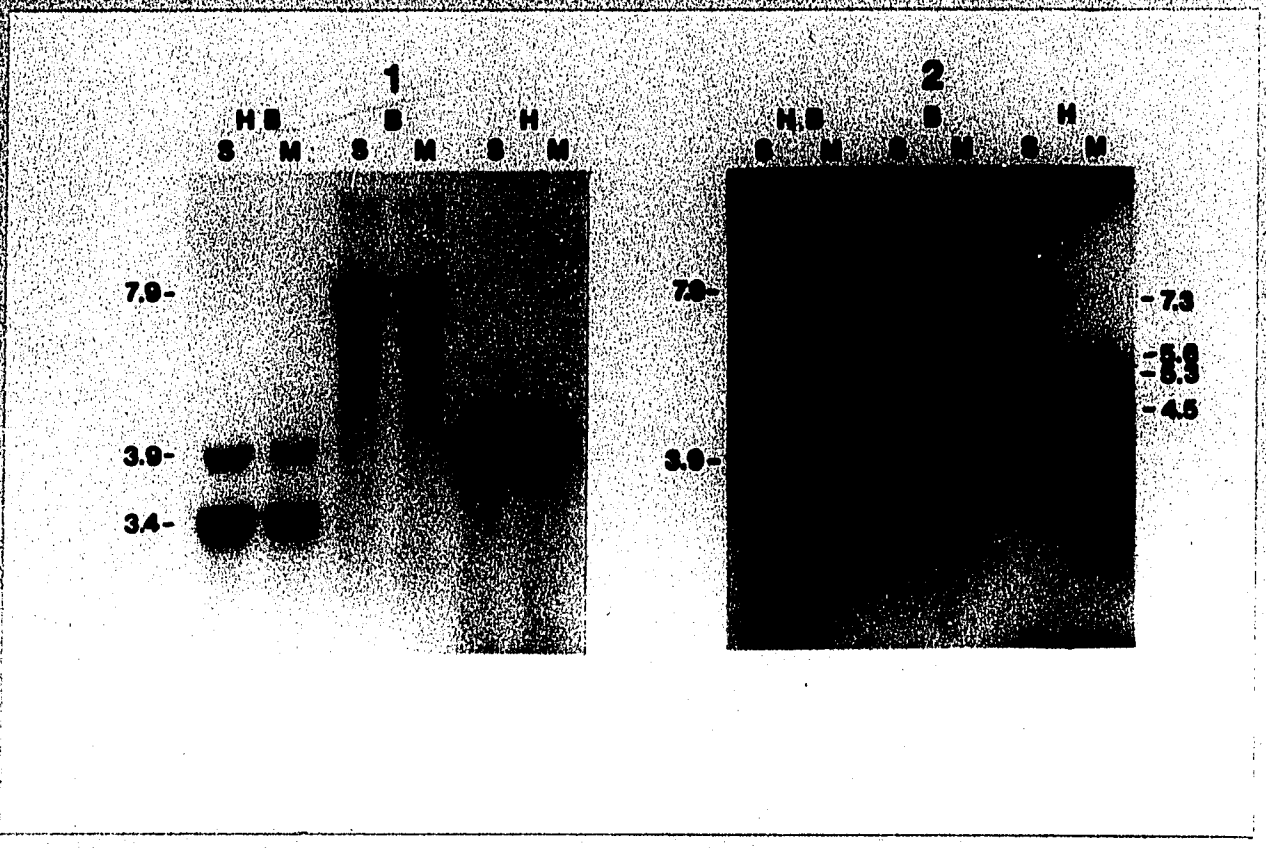
Figure 2. Analysis of the storage proteins of parents and G. soja PIs with a variant A_4 polypeptide using SDS-PAGE with 5 M urea in the gel. The electrophoretic conditions are given in the materials and methods section and the proteins are labelled as in Figure 1 with the exception of the basic polypeptides B_{1a}, B_{1b}, B_2 and B_4 being present (labelled B's). The lanes were loaded with the following: 1, Raiden; 2, A81-356022; 3, PI 468916; 4, molecular weight standards; 5, PI 407288; 6, PI 366122; 7, PI 339735; and 8, PI 468918


1
2
3

DraI, EcoRI, EcoRV, HindIII and TaqI to determine if RFLPs were present between the parents. All enzymes except BamHI revealed a RFLP between the parents. The location of the DNA rearrangement that resulted in the HindIII RFLP was determined by cutting pG-248 at an internal HindIII restriction site (Scallon et al., 1987) and using the 5' and 3' ends of the clone separately as hybridization probes (Figure 3). When HindIII-cut DNA of the parents was hybridized with the 5' end of the clone, a clear polymorphism was observed between a 4.5 kb fragment in G. max and a 7.3 kb fragment in G. soja (Figure 3). The 5.6 and 5.3 kb fragments were consistently less intense than the others and therefore were possibly the result of cross hybridization to Gy5 (Scallon et al., 1985). No polymorphism was observed when the 3' end was hybridized to HindIII-cut parental DNA (Figure 3). No polymorphism was observed by hybridizing either end of pG-248 to BamHI-cut DNA. These results suggested that the polymorphism was 5' from Gy4 and was located between the HindIH and the BamHI sites (Figure 4).

The 60 F_2 individuals in the population formed from a cross between the G. max and G. soja parents were genotyped both for the HindIII polymorphism with pG-248 and the protein variant using SDS-PAGE. The genotype of each F_2 was identical for both SDS-PAGE and RFLP analysis. This confirms that the variant polypeptide is A_4. It also suggests that the variant was the result of a modification to the Gy4 gene coding sequence and not from the action of a separate unlinked gene. The segregation data were analyzed using the program Mapmaker together with the data of over 300 markers scored in this

Figure 3. Autoradiographs of the hybridization of the 3' end (1) and 5' end (2) of pG-248 to HindIII (H), BamHI (B) and BamHI+HinDIII (HB) digested genomic DNA from the G. max parent, A81-356022 (M) and the G. soja parent, PI 468916 (S). Both autoradiographs were exposed to the same membrane. The membrane was first hybridized with the fragment from the 5' end, stripped, and then hybridized with the 3' end. The sizes of the fragments are given in kilobases


1
H B
S M
B
S M
H
S M
7.9-
3.9-
3.4-
2
H,B
S M
B
S M
H
S M
7.9-
3.0-
- 7.3
- 5.3
- 4.5

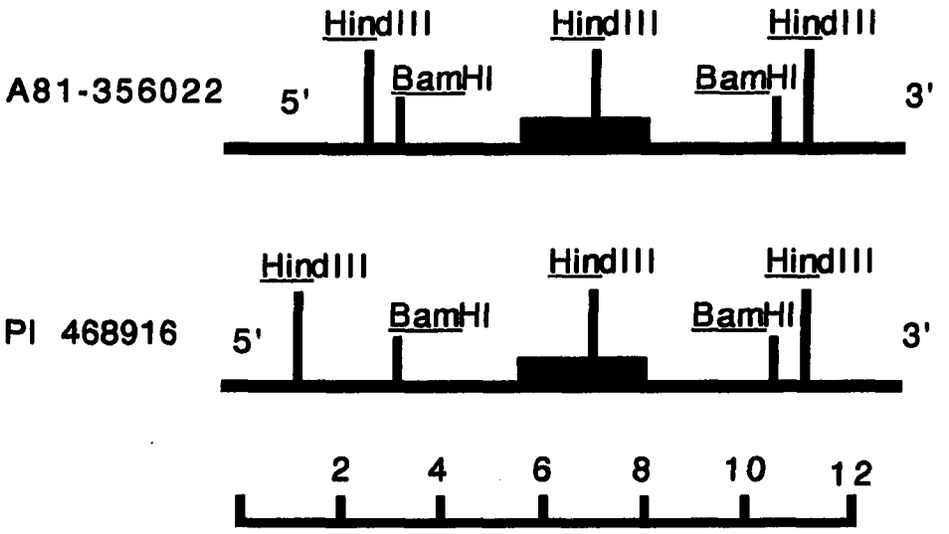


Figure 4. Restriction maps of Gy4 and surrounding regions for A81-356022 and PI 468916. The darkened box is the Gy4 gene. The scale is given in kilobases

population. These data were used to place Gy4 on the public RFLP map. (Fig. 5).

A possible cause of the A4 variant was an addition or insertion in the coding sequence of the A4 polypeptide in PI 468916. The DNA of the two parents were cut with several enzymes and hybridized with pG-248 to search for size differences in the coding sequence for A4. The size of the fragments obtained from the digestions were predicted from the published sequence information for Gy4 (Scallon et al., 1987). A polymorphic HinDIII+StyI fragment was identified. This fragment had a predicted size of 181 base pairs (bp) (Figure 6). The size difference between the fragments was consistent with the size variation of the A4 polypeptide because both the A4 polypeptide and HinDIII+StyI fragment from the G. soja parent had a lower mobility (higher molecular weight) than from the G. max parent. The HinDIII+StyI fragment includes 151 bp of the A4 coding sequence and 30 bp of the B3 coding sequence. Thus, it is not possible to determine unambiguously from these data whether the sequence change that resulted in the polymorphism is with the A4 or B3 coding sequence. However, because no mobility differences have been observed between the two parents for B3, it seems likely that the change is indeed within A4.

The carboxyl end of the acidic glycinins, where our data suggests that an addition or insertion in A4 is located, is referred to as the 'hypervariable region' (Nielsen et al., 1989). This region is highly

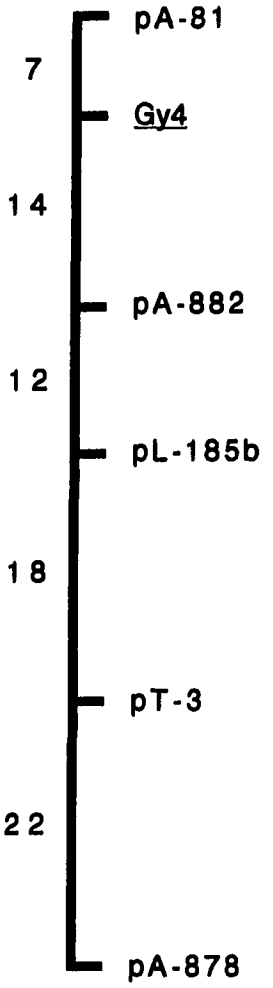
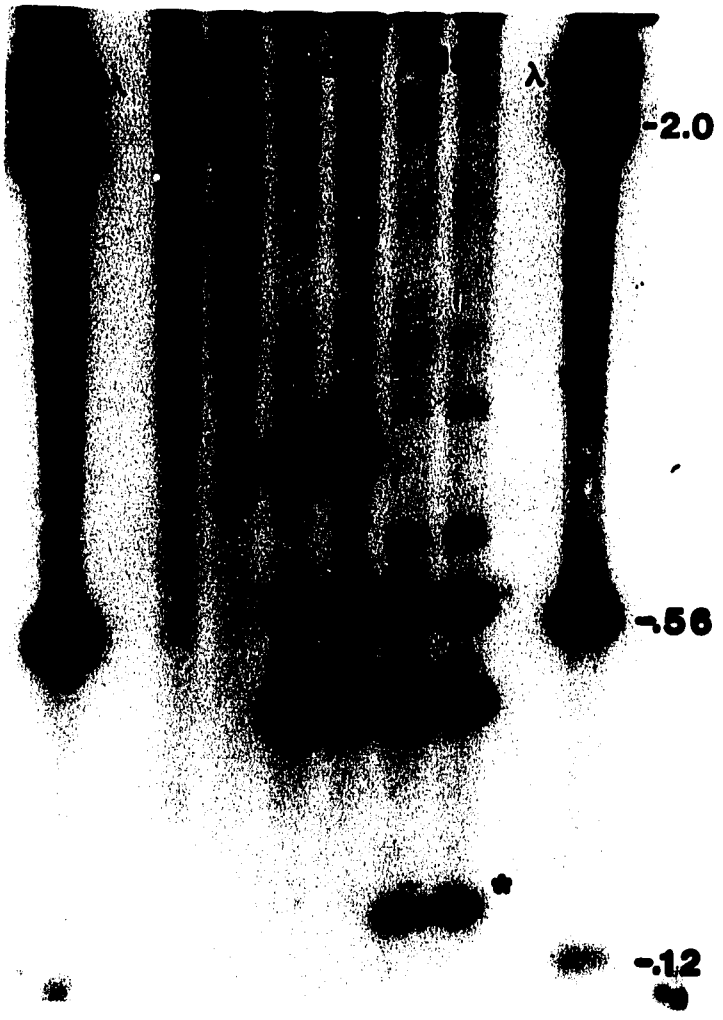


Figure 5. Linkage group that includes Gy4. Linkage arrangement and distances were determined using Mapmaker (Lander et al., 1987). The distances between markers are given in cMorgans to the left of the linkage group. Markers labelled pA were developed at Iowa State University and markers labelled pT and pL were developed at the University of Utah and kindly provided by Dr. K. G. Lark

Figure 6. Autoradiograph of the hybridization of the 5' end of pG-248 to HinDIII, StyI and HinDIII+StyI digested genomic DNA from the G. max parent, A81-356022 (M) and the G. soja parent, PI 468916 (S). Also included is HindIII cut λ DNA (λ). The sizes of the fragments are given in kilobases. A * designates the HinDIII+StyI fragment that is polymorphic


λ
-2.0
-.56
-.12

divergent among the glycinins and has a high content of the amino acids aspartate and glutamate. The region also has a sequence of 35 amino acids that are duplicated once in G5 and twice in G4. Certainly a possible molecular explanation of the A4 variant is that in PI 468916 there is a third duplication. Sequence analysis of this region would be required to determine this.

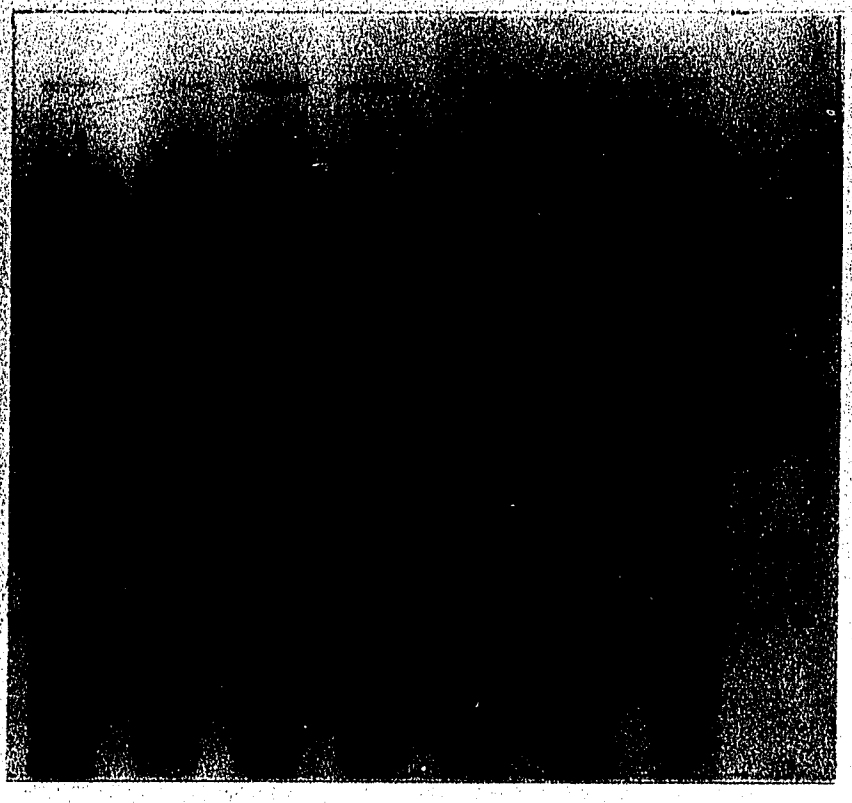
The survey of 24 G. soja accessions using SDS-PAGE revealed four more mobility variants for the A4 polypeptide (Table 1). Two of the accessions had A4 mobilities similar to PI 468916, the G. soja parent, and two had A4 mobilities that were faster than either parent (Figure 2, 7). The A4 polypeptides with fast mobilities were not separated from the other A's using standard gels but could be separated using urea gels (Figure 2). This variability found for A4 in G. soja was unexpected. Staswick et al. (1983) reported observing no mobility variants for storage proteins among 120 G. soja lines that they surveyed.

The accessions PI 468918 and PI 407288 had A4 mobilities similar to PI 468916, the G. soja parent of the mapping population. All three of the accessions were collected in North-eastern China. PI 468916 and PI 468918 were collected near Shenyang, in Liaoning Province and PI 407288 was collected near Kungchuling, in Jilin Province which is adjacent to Liaoning. The accessions PI 366122 and PI 339735A had another A4 variant. The A4 polypeptide in these had a faster mobility than either parent of the mapping population. PI 366122 was collected in Japan and PI 339735A was collected in Korea.

Table 1. List of the G. soja plant introductions screened for storage protein variation

PI Number	Country of origin
407293	China
407288	China
135624	China
101404	China
65549	China
468904	China
440913	China
468918	China
468906	China
468904	China
407299	China
407275	South Korea
407200	South Korea
407184	South Korea
407162	South Korea
339735	South Korea
424004	South Korea
366122	Japan
342618	Soviet Union
326582	Soviet Union
326581	Soviet Union
423991	Soviet Union
423988	Soviet Union

Figure 7. Analysis of storage proteins using SDS-PAGE. The electrophoretic conditions were given in the materials and methods section and the proteins are labelled as in Figure 1. The lanes were loaded with the following: 1, Raiden; 2, A81-356022; 3, PI 468916; 4, PI 407288; 5, PI 366122; 6, PI 339735; and 7, PI 468918



This type of mutants may provide a means by which the evolution of storage protein genes and gene families can be studied.

REFERENCES

Apuya, N., B. L. Frazier, P. Kiem, E. J. Roth and K. G. Lark. 1988. Restriction length polymorphisms as genetic markers in soybean, Glycine max (L.) Merr. Theor. Appl. Genet. 75:889-901.

Badley, R. A., D. Atkinson, H. Hauser, D. Odani, J. P. Green and J. M. Stubbs. 1975. The structure, physical and chemical properties of the soy bean protein glycinin. Biochim. Biophys. Acta 412:214-228.

Cho, T, C. S. Davies, R. L. Fischer, N. E. Tumer, R. B. Goldberg, and N. C. Nielsen. 1989. Molecular characterization of an aberrant allele for the Gy3 glycinin gene: a chromosomal rearrangement. The Plant Cell. 1:339-350.

Dickinson, C. D., L. A. Floener, G. G. Lilley, and N. C. Nielsen. 1987. Self-assembly of proglycinin and hybrid proglycinin synthesized in vitro from cDNA. Proc. Nat. Acad. Sci. 84:5525-5529.

Fischer, R. L., and R. B. Goldberg. 1982. Structure and flanking regions of soybean seed protein genes. Cell 29:651-660.

Fontes, B. P. B., M. A. Moreira, C. S. Davies, and N. C. Nielsen. 1984. Urea-elicited changes in relative electrophoretic mobility of certain glycinin and ß-conglycinin subunits. Plant Physiol. 76:840-842.

Honeycutt, R. J., J. W. Burton, R. G. Palmer, and R. C. Shoemaker. 1989. Association of major seed components with a shriveled-seed trait in soybean. Crop Sci. 29:804-809.

Keim, P., R. C. Shoemaker, and R. G. Palmer. 1989. Restriction fragment length polymorphism diversity in soybean. Theor. Appl. Genet. 77:786-792.

Keim, P., B. W. Diers, T. C. Olson, and R. C. Shoemaker. 1990. RFLP mapping in soybean: Association between marker loci and variation in quantitative traits. Genetics 126:735-742.

Kitamura, K., Y. Toyokawa, and K. Harada. 1980. Polymorphism of glycinin in soybean seeds. Phytochemistry 19:1841-1843.

Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) 227:680-685.

Lander E. S., P. Green, J. Abrahamson, A. Barlow, M. J. Daly, S. E. Lincoln, and L. Newburg. 1987. Mapmaker: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174-181.

Moreira, M. A., M. A. Hermodson, B. A. Larkins, and N. C. Nielsen. 1979. Partial characterization of the acidic and basic polypeptides of glycinin. J. Biol. Chem. 254:9921-9926.

Nielsen, N. C., C. D. Dickinson, T. Cho, V. H. Thanh, B. J. Scallon, R. L. Fischer, T. L. Sims, G. N. Drews, and R. B. Goldberg. 1989. Characterization of the glycinin gene family in soybean. Plant Cell 1:313-328.

Scallon, B., V. H. Thanh, L. A. Floener and N. C. Nielsen. 1985. Identification and characterization of DNA clones encoding group-II glycinin subunits. Theor. Appl. Genet. 70:510-519.

Scallon, B. J., Dickinson, C. D. and N. C. Nielsen. 1987. Characterization of a null-allele for the Gy4 glycinin gene from soybean. Mol. Gen. Genet. 208:107-113.

Staswick, P. E. and N. C. Nielsen. 1983. Characterization of a soybean cultivar lacking certain glycinin subunits. Arch. Biochem. Biophys. 223:1-8.

Staswick, P. E., P. Broue, and N. C. Nielsen. 1983. Glycinin composition of several perennial species related to soybean. Plant Physiol. 72:1114-1118.

Staswick, P. E., M. A. Hermodson and N. C. Nielsen. 1984. Identification of the cystines which link the acidic and basic components of the glycinin subunits. J. Biol. Chem. 259:13431-13435.

Tumer, N., J. D. Richter, and N. C. Nielsen. 1982. Structural characterization of the glycinin precursors. J. Biol. Chem. 257:4016-4018.

Wilson, R. F. 1987. Seed metabolism. In J. R. Wilcox (ed.) Soybeans: Improvement, production, and uses. 2nd ed. Agronomy 16:643-686.

GENERAL CONCLUSIONS

In the past, genetics in soybean has been less advanced than in many other important crop species. Recently, both a public and private RFLP map have been constructed in soybean (Keim et al., 1990b; Tingey et al., 1989). Saturated RFLP maps will greatly improve the level of genetic and molecular research that can be conducted in soybean. In the research presented in this thesis, markers from the public RFLP map were used to map both quantitative and qualitative traits. The mapping of qualitative genes and the genetic dissection of quantitative traits with RFLP markers should help to improve our understanding of the soybean and make RFLP maps more useful to breeders and geneticists.

The first study presented in this thesis is an extension of the mapping efforts reported by Keim et al. (1990a, 1990b). In this study, quantitative trait loci (QTLs) for seed size, protein, oil and relative fatty acid content were mapped. This mapping was conducted in the population being used to develop the public RFLP map, and which was used to map QTLs for hard seededness and vegetative growth traits (Keim et al., 1990a and 1990b). The seed component traits were mapped using the F_2 segregation data collected to form the RFLP map. Some highly significant associations between markers and the seed traits were found in this population. Two linkage groups strongly associated with seed traits were tested in a second population. The associations were consistent between

populations for one linkage group but not the other. These inconsistencies could have resulted from a number of causes, the most likely being genotype by environment interaction. Because of the importance of protein and oil in soybean, additional research will be initiated to map these traits in other populations. These studies should result in a better understanding of the genetic control of seed component traits that could help plant breeders.

In the second study, near isogenic lines (NILs) were used to map phytophthora resistance loci with RFLPs. There are six Rps loci that confer resistance to phytophthora root rot in soybean. Five of these resistance loci and a nodulation gene linked to a resistance locus were mapped. The markers that are linked to the Rps loci will be useful in further research. One area of research that is being pursued is to develop a system of marker based selection using allele specific oligonucleotides. Allele specific oligonucleotides could be designed based on the nucleotide sequences of markers that are linked to the resistance genes. Allele specific oligonucleotides in conjunction with polymerase chain reaction could give a fast and inexpensive method of selection for phytophthora resistance that is superior to the traditional methods. The markers also represent starting points for chromosome walking and the eventual cloning of the phytophthora resistance genes.

In the last study, the Gy4 gene that confers the glycinin subunit G4 was mapped and genetically analyzed. The study was initiated when a variant of the A4 polypeptide from the glycinin

storage protein was observed between the two parents being used to form the RFLP map. The A4 polypeptide in the G. soja parent had a lower mobility than the A4 polypeptide in the G. max parent, or in other any other genotype reported. Gy4 was then mapped based upon the segregation in the F_2 mapping population of the variant polypeptide and upon a RFLP revealed through hybridization of a clone of the Gy4 gene to HindIII-cut DNA of the F_2 individuals. Presently the molecular basis of the mutant causing the variant is being studied. Sequencing the region of A4 that encodes Gy4 should reveal this information.

ADDITIONAL LITERATURE CITED

Apuya, N. R., B. L. Frazier, P. Keim, E. J. Roth and K. G. Lark. 1988. Restriction fragment length polymorphisms as genetic markers in soybean, Glycine max (L.) Merrill. Theor. Appl. Genet. 75:889-901.

Beckmann, J. S., and M. Soller. 1983. Restriction fragment length polymorphisms in genetic improvement: Methodologies, mapping and costs. Theor. Appl. Genet. 67:35-43.

Bernard, R. L. 1976. United States national germplasm collections. p. 286-289. In L. D. Hill (ed.) World Soybean Research. Interstate Printers and Publ. Inc., Danville, IL.

Bubeck, D. M., M. M. Goodman, W. D. Beavis, and D. Grant. 1990. Quantitative trait loci controlling resistance to gray leaf spot in maize. p. 194. In Agronomy abstracts. ASA, Madison, WI.

Burr, B., S. V. Evola, F. A. Burr, J. S. Beckmann. 1983. The application of restriction fragment length polymorphism to plant breeding. pp. 45-59. In J. K. Setlow and A. Hollaender (eds.) Genetic engineering principles and methods. Vol. 5. Plenum Press, New York.

Burr, B. 1989. Introduction. In T. Helentjaris and B. Burr (ed.) Development and application of molecular markers to problems in plant genetics. pp. 1-3. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Edwards, M. D., C. W. Stuber, J. F. Wendel. 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, W. R. 1987. Breeding methods for cultivar development. In J. R. Wilcox (ed.) Soybeans: Improvement, production and uses. Agronomy 16:249-294.

Gardiner, J. M., D. A. Hoisington, and S. Melia-Hancock. 1989. Use of isogenic lines for mapping Ht loci in Zea mays L. p. 81. In Agronomy abstracts. ASA, Madison, WI.

Graef, G. L., W. R. Fehr, S. R. Cianzio. 1989. Relation of isozyme genotypes to quantitative characters in soybean. Crop Sci. 29:683-688.

Grant, D., B. Beavis, M. L. Katt, R. Feazel, R. Meier, R. Fincher and W. Nelson. 1989. Comparison of QTLs for maturity related traits among different crosses of maize. p. 176. In Agronomy abstracts. ASA, Madison, WI.

Hanson, W. D. 1959. Minimum family sizes for the planning of genetic experiments. Agron. J. 51:711-715.

Helentjaris, T., and D. Weber and S. Wright. 1988. Identification of the genomic locations of duplicate nucleotide sequences in maize by analysis of restriction fragment length polymorphisms. Genetics 118:353-363.

Hymowitz, and T., R. J. Singh. 1987. Taxonomy and speciation. In J. R. Wilcox (ed.) Soybean: Improvement, production and uses. 2nd ed. Agronomy 16:23-48.

Keim, P., R. C. Shoemaker, and R. G. Palmer. 1989. Restriction fragment length polymorphism diversity in soybean. Theor. Appl. Genet. 77:786-792.

Keim, P., B. W. Diers, R. C. Shoemaker. 1990a. Genetic analysis of soybean hard seededness with molecular markers. Theor. Appl. Genet. 79:465-469.

Keim, P., B. W. Diers, T. C. Olson, and R. C. Shoemaker. 1990b. RFLP mapping in soybean: Association between marker loci and variation in quantitative traits. Genetics. Accepted for publication.

Lander, E. S., D. Botstein. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185-199.

Martin, B., J. Nienhuis, G. King, and A. Schaefer. 1989. Restriction fragment length polymorphisms associated with water use efficiency in tomato. Science 243:1725-1728.

Muehlbauer, G. J., J. E. Secht, P. E. Staswick, G. L. Graef, and M. A. Thomas-Comptom. 1989. Application of the near-isogenic line gene mapping technique to isozyme markers. Crop Sci. 29:1548-1553.

Muehlbauer, G. J., J. E. Specht, M. A. Thomas-Compton, P. E. Staswick, and R. L. Bernard. 1988. Near-isogenic lines-A potential resource in the integration of conventional and molecular marker linkage maps. Crop Sci. 28:729-735.

Muehlbauer, G. J., P. E. Staswick, J. E. Specht, G. L. Graef, R. C. Shoemaker, and P. Keim. 1990. RFLP mapping in the soybean (Glycine max L. Merr.) using near-isogenic lines. Theor. Appl. Genet. (In review).

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

Osborn, T. C., D. C. Alexander, J. F. Fobes. 1987. Identification of restriction fragment length polymorphisms linked to genes controlling soluble solids content in tomato fruit. Theor. Appl. Genet. 73:350-356.

Palmer, R. G., T. C. Kilen. 1987. Qualitative genetics and cytogenetics. In J. R. Wilcox (ed.) Soybean: Improvement, production and uses. 2nd ed. Agronomy 16:23-48.

Paterson, A. H., E. S. Lander, J. D. Hewitt, S. Peterson, S. E. Lincoln, S. D. Tanksley. 1988. Resolution of quantitative traits into mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721-726.

Rafalski, J. A. and S. Tingey. 1990. Genetic mapping of biochemically defined loci in soybean using cloned gene probes. p. 24. In Proceedings of the 3rd biennial conference on molecular and cellular biology of the soybean. Iowa State University, Ames, IA.

Sax, K. 1923. The association of size differences with seed coat pattern and pigmentation in Phaseolous vulgarus. Genetics 8:552-560.

Smith, K. J. and W. Huyser. 1987. World distribution and significance of soybean. In J. R. Wilcox (ed.) Soybeans: Improvement, production and uses. Agronomy 16:1-22.

Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503-517.

Suarez, J. C. 1989. Association between isozyme genotypes with quantitative traits in soybean. Ph.D. diss. Iowa State Univ., Ames (Diss. Abstr. 89-20190)

Tanksley, S. D. and J. Hewitt. 1988. Use of molecular markers in breeding for soluble solids content in tomato-a reexamination. Theor. Appl. Genet. 75:811-823.

Tanksley, S. D., H. Medina-Filho, and C. M. Rick. 1982. Use of naturally-occurring enzyme variation to detect and map genes controlling quantitative traits in an interspecific backcross of tomato. Heredity 49:11-25.

Tingey, S., S. Sebastian, and J. A. Rafalski. 1989. A RFLP map of the soybean genome. p. 180. In Agronomy abstracts. ASA, Madison, WI.

Webb, D. M., S. J. Knapp and L. A. Tagliani. 1990. An RFLP and allozyme linkage map of Cuphea lanceolata and QTL affecting seed fatty acids, seed oil, and embryo weight. p. 204. In Agronomy abstracts. ASA, Madison, WI.

Young, N. D., D. Zamir, M. W. Ganal and S. D. Tanksley. 1988. Use of isogenic lines and simultaneous probing to identify DNA markers tightly linked to the Tm-2a gene in tomato. Genetics 120:579-585.

ACKNOWLEDGEMENTS

"An education is the going forth from cocksure ignorance to thoughtful uncertainty." Unknown.

Had I been told when I started as an undergraduate that in ten years time I would still be a student at university, I would have found it unbelievable. But life unfolds in unexpected ways and here I am with very few regrets. There are many people that I must acknowledge for contributing to my degree and below is certainly an incomplete list.

I would first like to acknowledge Dr. Randy Shoemaker, my advisor and who I would like to think as a friend. His unfailing faith in my ideas and work was a great encouragement. I appreciate his patients during the writing of this thesis. Thanks also go to Dr. W. R. Fehr who encouraged me to stay at Iowa State for my Ph. D. and gave me a fellowship that has supported me during this degree.

I also thank my other committee members, Drs. Thornburg, Wendel and Palmer with whom I have conferred many times during the course of my study and have always been helpful. I am indebted to Paul Keim whose encyclopedic knowledge of science and technical skill in the lab are astounding. I must credit him for teaching me the basic skills in molecular biology and making me excited about practicing them.

Thanks go to the members of Randy's lab whose support and friendship I appreciated. Certainly the relentless efforts of Terry Olson in developing and mapping markers have made my work much easier.

I would like to recognize some of my research collaborators. In the phytophthora mapping study, the phytophthora screening was done in the lab of Dr. Levi Mansur with expertise of Rick Ruff. The nodulation screening was set up with the help of Dr. John Imsande and the rhizobium cultures were provided by Dr. Renee Kosslak. The work in the protein variant study was expedited by receiving the cDNA clone of Gy4 from Dr. Niels Nielsen, Purdue University. I also would like to thank Dr. Reid Palmer for supplying me with the G. soja PIs that I screened in the protein variant study and for his careful reading of this thesis.

Thanks also go to my friends in Ames who have made my study in this town enjoyable. Christer Ohlsson, Ramon Gatnau and Peter Lynch have been good housemates and friends during my stay. Our 6:00 dinner, discussion and coffee that we have had together almost every weekday these past years has been a major factor in keeping my sanity.

Finally I must credit my parents, Gordon and Lois for their understanding and support during these years of study. I am sure by now you are tired of explaining to the relatives and neighbors why Brian still doesn't have a real job.