MICROSATELLITE POLYMORPHISM, ORTHOLOGOUS EVOLUTION AND MOLECULAR MARKER ANALYSIS OF SEED QUALITY TRAITS IN SOYBEAN (*Glycine max* L. Merr.)

by

PETER JEFFREY MAUGHAN

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

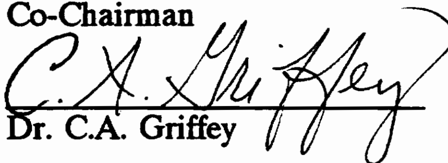
GENETICS

Approved:

Dr. G.R. Buss
Co-Chairman

Dr. M.A. Saghai Maroof
Co-Chairman

Dr. A. Esen



Dr. C.A. Griffey

Dr. C.A. Wilkinson

December, 1994
Blacksburg, Virginia

c.2

LD
5655
V856
1994
M5865
c.2

Microsatellite polymorphism, orthologous evolution and molecular marker analysis of seed quality traits in soybean (Glycine max L. Merr.)

by

Peter Jeffrey Maughan

G.R. Buss, Co-Chairman

M.A. Saghai Maroof, Co-Chairman

Genetics

(ABSTRACT)

In this study we assayed the extent of genetic variation for five microsatellites in 94 accessions of wild (*Glycine soja*) and cultivated soybean (*G. max*). F_2 segregation analysis indicated that all five of the microsatellites were independently inherited and four loci were located in four independent linkage groups. The number of alleles per microsatellite locus ranged from five to 21. Overall, 43 more microsatellite alleles were detected in wild than in cultivated soybean. Allelic diversity for microsatellite loci was significantly higher in wild than in cultivated soybean.

In a second study, molecular markers were used to identify and characterize quantitative trait loci (QTL) controlling seed weight in soybean, and to extend reports of orthologous seed weight genes in the genus *Vigna* to the genus *Glycine* by "comparative QTL mapping". DNA samples from 150 F_2 individuals from an interspecific soybean cross were analyzed with 91 genetic markers. Three and five markers were significantly associated with seed weight variation ($P<0.01$) in the F_2

and $F_{2:3}$ generations, respectively. Two-way ANOVA tests for digenic interactions identified three significant epistatic interactions in both generations. In a combined analysis, the significant marker loci and epistatic interactions explained 50 and 60% of the total variation for seed weight in the F_2 and $F_{2:3}$ generations, respectively. Comparison of our results in *Glycine* with those reported in *Vigna* indicated that both genera share orthologous seed weight genes. Moreover, a significant epistatic interaction between seed weight QTLs was conserved in both genera.

The objective of the third study was to use molecular markers and interval mapping techniques to position and characterize quantitative trait loci controlling seed protein, oil, sucrose, and calcium content as well as seed weight in soybean. Two QTLs were detected for protein and calcium content, five for oil content and seed weight and six for sucrose content, respectively. Percent phenotypic variation explained by these individual QTLs ranged from 6.6 to 34.0%. The total phenotypic variation explained by all QTLs for specific traits were 42.5%, 36.7%, 49.0%, 53.1%, and 42.6% for seed weight, protein, oil, sucrose, and calcium, respectively. Of the 11 genomic intervals identified in this study, six were associated with more than one seed quality trait. These results suggest that the genetic correlations observed between seed quality traits may be due to a pleiotropic effect of a single QTL or that QTLs controlling different seed quality traits were inherited in clusters as tightly linked loci.

ACKNOWLEDGEMENT

I express my appreciation to my advisors, Drs. G.R. Buss and M.A. Saghai Maroof, for their endless encouragement, guidance, and patience throughout my studies and during the preparation of this dissertation. I would also like to acknowledge the guidance and assistance received from the other members of my advisory committee: Dr. C.A. Griffey, Dr. C.A. Wilkinson, and Dr. A. Esen. I am honored to have associated with such outstanding scientists.

I also would like to express my sincere gratitude to those students, research scientist, and visiting scholars with whom I have enjoyed friendships and collaborations, namely: Y.G. Yu, G. Ma, P. Bagade, C. Ndirangu, J. Zhang, D. Ball, Dr. R. Biyashev, Dr. G. Yang, Dr. P. Chen, Dr. R. Ragab, Dr. S. Dranovalli, and Dr. Q. Zhang.

I express my sincere love for my caring parents, Peter and Judy Maughan, whose support and encouragement have made possible my education. Lastly I give my greatest thanks to my beautiful wife, Dawn, whose love, patience, and understanding during the last four years have made my education and scientific endeavors feasible.

TABLE OF CONTENTS

	Pages
ABSTRACT	ii
ACKNOWLEDGEMENT	iv
TABLE OF CONTENTS	v
LIST OF TABLES	viii
LIST OF FIGURES	ix
CHAPTER	
I. INTRODUCTION	1
Restriction fragment length polymorphism	3
Molecular markers and plant breeding	4
Interval mapping	6
Molecular marker analysis in soybean	7
Microsatellites	9
Comparative QTL mapping and orthologous evolution	12
Objectives of this study	14
References	15
II. MICROSATELLITE AND AMPLIFIED SEQUENCE LENGTH POLYMORPHISMS IN CULTIVATED AND WILD SOYBEAN	23
Abstract	24

Introduction 25

Materials and Methods 27

Results 30

Discussion 35

References 40

Tables and Figures 43

III. MOLECULAR MARKER ANALYSIS OF SEED WEIGHT: GENOMIC LOCATIONS, GENE ACTION, AND EVIDENCE FOR ORTHOLOGOUS EVOLUTION AMONG THREE LEGUME SPECIES 52

Abstract 53

Introduction 55

Materials and Methods 57

Results 61

Discussion 65

References 70

Tables and Figures 73

VI. GENETIC ANALYSIS OF QUANTITATIVE TRAIT LOCI INFLUENCING SEED QUALITY CHARACTERISTICS IN SOYBEAN USING MOLECULAR MARKERS AND INTERVAL MAPPING TECHNIQUES 81
Abstract 82
Introduction 83
Materials and Methods 85
Results 89
Discussion 92
References 97
Tables and Figures 100

VITA 106

LIST OF TABLES

CHAPTER II.

Table 1. Oligonucleotide primer sequences designed to detect SSR and ASLP variation in soybean . . . 49

Table 2. Number of alleles, level of diversity, and differentiation in *Glycine max* and *Glycine soja* detected at four SSR and five ASLP loci . . . 50

Table 3. Frequencies of the two most prevalent alleles of each SSR and ASLP locus . . . 51

CHAPTER III.

Table 1. Location, gene action, and effect of markers significantly associated with seed weight . . . 79

Table 2. Significant interactions between genetic markers associated with seed size as determined using two-way ANOVA . . . 80

CHAPTER IV.

Table 1. Means, standard deviations, ranges and parental values for five seed quality traits . . . 103

Table 2. Pearson correlation coefficients among five seed quality traits measured in $F_{2:3}$ families . . . 104

Table 3. Location, gene action, and effect of significant intervals containing quantitative trait loci (QTL) for seed quality characteristics . . . 105

LIST OF FIGURES

CHAPTER II.

Figure 1. Microsatellite polymorphism at locus PRP1 43

Figure 2. Continuous allelic ladders for two soybean microsatellite loci . . . 44

Figure 3. Continuous allelic ladder for the soybean microsatellite locus HSP176L 45

Figure 4. Discontinuous allelic ladder for the soybean microsatellite HSP179D 46

Figure 5. F_2 segregation pattern for the microsatellite HSP179D. 47

Figure 6. DNA polymorphism at three ASLP loci 48

CHAPTER III.

Figure 1. Frequency distribution for seed weight in the F_2 and $F_{2:3}$ generations 73

Figure 2. Regression plot of F_2 seed weight on $F_{2:3}$ seed weight 74

Figure 3. Regression plot showing the association of seed weight with percent heterozygous loci 75

Figure 4. Orthologous genomic regions influencing seed weight in soybean, cowpea, and mungbean 76

Figure 5. Epistatic interaction between marker loci sgA816 and mgM185b in the F_2 and $F_{2:3}$ generations 77

Figure 6. Genotypic composition of the F_2 population 78

CHAPTER IV.

Figure 1. Frequency distributions for seed weight, protein, oil, sucrose, calcium, and ln(calcium) 100

Figure 2. QTL likelihood plot for a genomic region influencing multiple seed quality traits in soybean 102

CHAPTER I.

Introduction

The science of plant genetics traces to classic studies on monogenic inheritance of morphological traits performed by Mendel (Mendel 1865). Since that time plant geneticists have been identifying, cataloging and mapping single gene markers in many plant species. Indeed, genetic maps for plant species were among the first to be constructed (Emerson et al. 1935; MacArthur 1934). Until recently, the source of genetic markers for such maps have been mutations which affect plant morphology (e.g., dwarfism).

In 1923, Sax proposed the use of genetic markers linked to genes of interest as a selection method for the improvement of complex agronomic traits. Subsequently, several researchers reported linkages between single genetic markers and quantitative trait loci (QTLs) (Rasmusson 1935; Everson and Schaller 1955; Thoday 1961). Unfortunately, the morphological markers used in these classic studies posed major limitations to such research because i) only a few markers could be studied in a single cross, ii) these mutations were often deleterious, iii) these markers were subject to large environmental effects, and iv) such markers often masked phenotypic effects of linked genes, making it nearly impossible to identify desirable linkages for selection. The advent of isozyme markers ameliorated some of the potential problems encountered by early plant geneticists. However, the number of isozyme markers available for genetic analysis is limited and insufficient for many of the potential applications of genetic mapping in plant breeding (Tanksley and Orton 1983).

Restriction fragment length polymorphism

Restriction fragment length polymorphisms (RFLPs) result from changes in DNA sequences. These changes may result from point mutations, deletions, insertions, or transposition events. The consequence of such changes results in DNA fragments of differing sizes from individuals with different genotypes. Different sized DNA fragments, separated by agarose gel electrophoresis, can be visualized by autoradiography. Segregation analysis of multiple RFLP markers can then be used to produce highly saturated genetic maps. RFLP markers have several advantages over conventional genetic markers (e.g., morphological or isozyme markers) in that they are i) codominant and highly informative (heterozygotes can be distinguished from either homozygote), ii) unaffected by the environment, and iii) abundant and highly polymorphic in most species.

RFLPs were first described for adenoviruses (Grodziker et al. 1974) and later for humans (Jeffreys 1979; Botstein et al. 1980). Botstein et al. (1980) first proposed the use of RFLPs as genetic markers for linkage map construction in humans. Early reports of linkage of RFLPs to heritable monogenic disease loci included Huntington's disease (Gusella et al. 1983), ß-thalassemia (Little et al. 1980), and sickle cell anemia (Phillips et al. 1980). Donis-Keller et al. (1987) published the first complete genetic linkage map of the human genome, based primarily on RFLP markers.

Although the use of RFLPs in plant systems was examined in theoretical terms by several authors (Burr et al. 1983; Soller and Beckmann 1983; Tanksley et al 1983), early investigations were more limited, perhaps because plant breeders and geneticists were unfamiliar with the required techniques. Helentjaris et al. (1985) reported the first applied study of RFLPs in plant studies. They found a high level of variability, as detected by RFLP markers, in maize cultivars and low levels of variability among cultivars of domesticated tomato. Early RFLP linkage maps constructed in plants included tomato (Bernatzky and Tanksley 1986), maize (Helentjaris et al. 1987), lettuce (Landry 1987), potato (Bonierbale et al. 1988), *Arabidopsis thaliana* (Chang et al. 1988), and rice (McCouch et al. 1988).

Molecular markers and plant breeding

The utility of molecular markers in plant breeding is based on the detection of RFLP markers which are tightly linked to genes of interest. Such linkage allows the breeder to infer the presence of a desirable gene by assaying directly for the RFLP marker. Often breeders are interested in transferring such genes into new genetic backgrounds. Traditionally, progeny are screened for the presence of the desired genes via phenotypic selection. However, simultaneous or even sequential screening of plants for several monogenic traits (e.g., several disease resistance genes) can be difficult or nearly impossible, especially when related genes can not be distinguished

by conventional methods. In contrast, linkages between multiple genes of interest and molecular markers make it practical to screen for many traits simultaneously without the need for costly selection procedures. The use of molecular markers as an aid to breeding has been coined "marker assisted selection" (MAS) by Lande and Thompson (1990). The first plant traits (genes) linked to RFLP markers included resistance to downy mildew in lettuce (Landry et al. 1987) and tobacco mosaic virus in tomato (Young et al. 1988) . No reports of marker assisted selection in conventional breeding have been published.

While some plant characters are controlled by single genes, the majority of agronomically important traits are controlled by the joint action of multiple genes. Such traits are said to be controlled in a polygenic or quantitative fashion. Until recently, plant breeders have relied on complex selection methods to improve specific polygenic traits. Such methods are expensive, time consuming, labor intensive, and require extensive resources. While breeders have been successful in manipulating many of these quantitative traits, little is known about the chromosomal location, number, or action of the genes controlling these traits. If polygenic traits, such as yield, could be resolved into individual genes, breeders might be able to deal with these complex traits with the same efficiency as with traits controlled by single genes. High density genetic mapping with molecular markers provides the means to identify and quantify genetic components of complex polygenic traits. Once individual genes are identified, selection can be based on the presence of a particular genetic

component, in much the same fashion as one would deal with a single gene trait. The first studies using RFLPs to map quantitative trait loci in plants were reported by Nienhuis et al. (1987) and Osborn et al. (1987), who characterized QTLs in tomato for insect resistance and fruit characteristics, respectively.

Interval mapping

The ability to detect a QTL with an RFLP marker is a function of the magnitude of the QTL's effect on the phenotype and the distance between the marker and the QTL. The traditional approach for detecting a QTL linked to a genetic marker involves a comparison of genotypic class means using one-way analysis of variance or linear regression (Edwards et al. 1987; Keim et al. 1990b). Moreover the coefficient of determination (r^2) provides an estimate of the genotypic effect of the QTL. Although the ANOVA approach provides a rough estimate of the position and effect of a QTL, it has several shortcomings: i) if the marker locus is not tightly linked to the QTL, estimates of phenotypic effects may be seriously underestimated, ii) the approach does not define the percise position of the QTL, but merely provides a general location, and iii) the suggested false positive rate of $\alpha=0.05$ for each ANOVA test leads to a large probability that at least one false positive will occur somewhere in the genome, especially if many markers are being tested.

To overcome these problems, Lander and Botstein (1989) suggested the use of interval mapping. This approach is based on two linked markers flanking an interval containing a QTL. Among the advantages of this approach are i) the inferred phenotypic effects are unbiased, because interval mapping qualifies the strength of the evidence for the QTL at multiple points within the interval containing the QTL, ii) the probable position of the QTL, in reference to the flanking markers, is predicted by the model, iii) interval mapping requires fewer progeny, because the genotype of the QTL can be inferred by the flanking markers, and iv) interval mapping can be used to distinguish a pair of linked QTLs from a single QTL. Interval mapping of QTLs has been utilized in the study of a variety of quantitatively controlled traits, including resistance to bacterial wilt in tomato (Danesh et al. 1994), cyst-nematode resistance in potato (Kreike et al. 1993), testcross performance in maize (Schön et al. 1994; Stuber et al. 1992), soybean cyst nematode resistance (Concibido et al. 1994), and reproductive and morphological traits in soybean (Mansur et al. 1993).

Molecular marker analysis in soybean

In soybean, the level of genetic polymorphism for RFLPs was first investigated by Apuya et al. (1988) and Doyle (1988). The level of polymorphism detected by these researchers was extremely low and prevented extensive genetic mapping. A

survey of 58 wild and cultivated soybean accessions, however, identified genetically diverse genotypes suitable for RFLP map construction (Kiem et al. 1989). Kiem et al. (1990a) reported the first RFLP genetic map for soybean. This map was developed from an interspecific cross of a *G. max* breeding line and a *G. soja* plant introduction and consisted of 150 RFLP markers which identified 26 linkage groups spanning 1200 recombination units. The first analysis of quantitative traits in soybean using RFLPs also was reported by Keim et al. (1990a). Using one-way analysis of variance and genotypic class means, they found significant associations between molecular markers and leaf width ($r^2=0.24$), stem diameter ($r^2=0.24$), canopy height ($r^2=0.20$), and first flower ($r^2=0.23$). Keim et al. (1990b) also reported the association of five independent RFLP markers with variation in hard seededness in soybean. These markers (or nearby QTLs) and the epistatic interactions between them explained 71% of the total phenotypic variation for hard seededness. Subsequently, many genes controlling agronomically important traits in soybean have been mapped with RFLPs, including genes controlling supernodulation (Landau-Ellis et al., 1991) and *Phytophthora* root rot resistance (Diers et al., 1992a).

Diers et al. (1992b) expanded the soybean RFLP map to a total of 243 markers, spanning 2,147 centiMorgans (cM), and identified QTLs for seed protein and oil content. Keim et al. (1992) used 128 RFLP markers to evaluate the level of genetic diversity in U.S. soybean breeding populations. Shoemaker et al. (1992) conducted studies to characterize the genetics and genomic structure of soybean as a potential

ancient tetraploid. In 1993, Lark et al. developed a RFLP map of soybean using an intraspecific cross between the cultivars 'Minsoy' and 'Noir 1'. This map consisted of 132 RFLP, isozyme, morphological, and biochemical markers. The map defined 1550 cM and covered 31 linkage groups. Mansur et al. (1993) utilized this intraspecific cross and map to identify QTLs for 15 traits including reproductive, morphological, and seed traits. Most recently, DNA markers have been used to analyze resistance to soybean cyst nematode (Concibido et al. 1994) and soybean mosaic virus (Yu et al. 1994).

Microsatellites

Notwithstanding the great utility of RFLPs, they still pose several major limitations for researchers, including i) the rate of RFLP polymorphism is low in several plant species, especially among inbreeding species, including soybean (Helentjaris et al. 1985; Apuya et al. 1988), ii) identification and characterization of RFLPs is time consuming and difficult to automate, and iii) the availability of RFLP clones involves managing and maintaining large numbers of bacterial cultures. The first concern is especially serious. Although interspecific crosses often ensure phenotypic and molecular diversity, many traits of agronomic importance, such as yield, cannot be meaningfully evaluated in an interspecific cross. This is because plants from interspecific crosses often exhibit weedy characteristics, such as lodging

and shattering. Ideally, crosses should be designed according to phenotypic and biological considerations, rather than to maximize polymorphism for mapping.

Microsatellites (Litt and Luty 1989) or simple sequence repeats (Tautz 1989; Weber and May 1989) have recently been described as an alternative source of DNA polymorphism. Microsatellite loci consist of short tandemly repeated core sequences. The repeated core sequences, usually two to four nucleotides in length, often vary in number and are flanked by conserved DNA sequences. Thus, primers complementary to the flanking regions can be used to amplify the locus via the polymerase chain reaction (PCR). The PCR reaction includes one ^{32}P-labelled nucleotide to allow visualization of the amplified product via autoradiography after polyacrylamide gel electrophoresis. Microsatellites are ideal genetic markers in that they are i) abundant and appear to be evenly distributed throughout the genome (Weber 1990), ii) highly polymorphic (Litt and Lutty 1989), iii) rapidly typed via PCR, vi) codominant, and v) very accessible to other laboratories via published primer sequences.

Dinucleotide repeats are the most abundant type of microsatellite in mammalian genomes. The dinucleotide repeat $(GT)_n$ is estimated to occur at a frequency of 50,000 to 100,000 times or once every 30 to 60 kilobases (kb) in the mammalian genome (Hamada et al. 1982; Starling et al. 1991). Microsatellites of tri- and tetranucleotide repeats are also present at high frequencies (Tautz 1989) in the human genome. Genetic linkage maps based solely on microsatellites have been

constructed for a number of mammalian species, including humans (Weissenbach et al. 1992), mouse (Dietrich et al. 1992), rat (Serikawa et al. 1992), and cattle (Bishop et al. 1994).

While microsatellite variation has been increasingly exploited in mammalian systems, plant microsatellites remain virtually unexamined. Condit and Hubbell (1990) examined the abundance of dinucleotide repeats in maize and five tropical tree genomes and estimated 5 X 10^3 to 3 X 10^5 $(AC)_n$ and $(AG)_n$ sites per genome. Zhao and Kochert (1992) identified and characterized a $(GGC)_n$ microsatellite in rice. Akkaya et al. (1992) examined the degree of polymorphism detected by two $(AT)_n$ and one $(ATT)_n$ microsatellite loci in 43 soybean accessions. Each locus examined identified between six to eight allelic variants. Morgante and Olivieri (1993) found seven and eight alleles for two microsatellites in seven and 10 unrelated soybean lines, respectively. Studying four microsatellites in 15 plants of the genus *Brassica*, Langercrantz et al. (1993) found a mean of 4.2 alleles per microsatellite. Using eight inbred maize plants, Senior and Huen (1993) analyzed and described the inheritance of five maize microsatellite loci. These microsatellites were inherited in a Mendelian fashion and exhibited a polymorphic rate of 3.5 alleles per locus. Wu and Tanksley (1993) observed from five to 11 alleles at eight microsatellite loci in a sample of 20 rice accessions.

More recently, Saghai Maroof et al. (1994) reported the inheritance, polymorphism, and population dynamics of four barley microsatellite loci in a sample

of 207 wild and cultivated barley accessions. They found that the number of alleles ranged from three to 37 and that allelic diversity for microsatellite loci was significantly higher in wild than cultivated barley. Yang et al. (1994) examined genetic variation at 10 microsatellite loci in 238 accessions of landraces and cultivars of rice. They report the number of alleles varied from three to 25 and that about two-thirds as many alleles were observed in cultivars as in landraces of rice. Yu et al. (1994) reported the first genetic mapping of a disease resistance gene in plants using a soybean microsatellite marker, and Zhang et al. (1994a,b) used microsatellites to identify chromosomal regions in rice that have significant effects on grain yield.

Comparative QTL Mapping and Orthologous Evolution

Many of the problems faced by plant breeders can be ameliorated by the use of molecular markers and genetic mapping techniques. In addition to providing plant breeders with the ability to deal with complex agronomic traits as single Mendelian loci, molecular markers have applications in i) heterotic grouping of lines (Dudley et al. 1992), ii) positional cloning (Martin et al. 1993), iii) phylogenetic studies (Wang et al. 1992), and iv) comparative genome analysis (Bonierbale et al. 1988).

Comparative genome analysis is of particular interest to many quantitative geneticists. This analysis involves the use of molecular markers such as RFLPs or

microsatellites to map the genomes of two species for a common set of marker loci. Although this is labor intensive and expensive, comparative genome mapping allows one to determine the order of homologous DNA sequences along the chromosomes of distantly related species. Knowledge of homologous sequences among divergent species not only adds insight into chromosome evolution and speciation but also may provide a basis for interpreting genetic information among species. For example, mice are now use as a popular genetic model for the study of human genetic disorders, in part because mice and humans share several conserved linkage blocks (Davisson et al 1991; Darling and Abbott 1992).

Extensive conservation of linkage relationships have been reported between tomato and potato (Bonierbale et al. 1988), sorghum and maize (Whitkus et al. 1992), pea and lentil (Weeden et al. 1992), rice and maize (Ahn and Tanksley 1993), cowpea and mung bean (Menancio-Hautea 1993), and rice and wheat (Kurata et al. 1994). This conservation in linkage order among plant species has prompted some researchers to suggest that DNA markers tightly linked to genes of importance in one species may also be linked to the orthologous genes in related species (Paterson et al. 1991). Orthologous genes are genes in different species that diverged from a common ancestral gene (Fitch 1977). The use of QTL mapping information from one species to predict QTL properties in a related species has been termed "comparative QTL mapping" by Paterson et al. (1991). Recently, Fatokun et al. (1992) provided RFLP evidence for conservation of orthologous genes for a

quantitative trait among two species of the genus *Vigna*. In both cowpea and mung bean the genomic region with the greatest effect on seed weight spanned the same RFLP markers in the same linkage order. Whether "comparative QTL mapping" will be a valuable genetic tool for quantitative trait analysis across genera has yet to be investigated. Although soybean belongs to the genus *Glycine* while cowpea and mung bean belong to genus *Vigna*, all three species are from the same tribe (Phaseoleae). Thus, it may be possible that these genera share orthologous genes.

Objectives of this Study

The primary objectives of this research were three-fold: i) assess the usefulness of microsatellites as genetic markers in soybean, ii) identify and characterize quantitative trait loci controlling the expression of seed quality traits in soybean using molecular markers (RFLPs, SSRs, and RAPDs), and iii) use "comparative QTL mapping" techniques to determine whether soybean shares orthologous seed weight genes with cowpea and/or mung bean.

References

Ahn S, SD Tanksley (1993) Comparative linkage maps of the rice and maize genomes. Proc Natl Acad Sci USA 90:7980-7984

Akkaya MS, AA Bhagwat, PB Cregan (1992) Length polymorphisms of simple seqence repeat DNA in soybean. Genetics 132:1131-1139

Apuya NR, BL Frazier, P Keim, E Jill Roth, KG Lark (1988) Restriction fragment length polymorphisms as genetic markers in soybean, *Glycine max* (L.) Merill. Theor Appl Genet 75:889-901

Bernatzky R , SD Tanksley (1986) Toward a saturated linkage map in tomato based on isozymes and random cDNA sequences. Genetics 112:887-898

Bishop MD, SM Kappes, JW Keele, RT Stone, SLF Sunden, GA Hawkins, S Solinas Toldo, R Fries, MD Grosz, J Yoo, EW Beattie (1994) A genetic linkage map for cattle. Genetics 136:619-639

Bonierbale M, R Plaisted, SD Tanksley (1988) RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics 120:1095-1103

Botstein D, RL White, M Skolnick, RW Davis (1980) Construction of a genetic linkage map in man using reestriction fragment length polymorphisms. Am J. Hum. Genet. 32:314-331

Burr B, Evola E, Burr FA, Bechman JS (1983) The application of restriction fragment length polymorphism to plant breeding. In: Setlow JK, A Hollaender (eds) Genetic engineering principles and methods. Plenum Press, New York

Chang C, JC bowman, AW DeJohn, ES Lander, ES Meyerowitz (1988) Restriction fragment length polymorphism linkage map for *Arabidopsis thaliana*. Proc Natl Acad Sci USA 85:6856-6860

Concibido V, RL Denny, SR Boutin, R Hautea, JH Orf, ND Young (1994) DNA marker analysis of loci underlying resistance to soybean cyst nematode (*Heterodera glycines* Ichinoche). Crop Sci 34:240-246

Condit R, SP Hubbell (1991) Abundance and DNA sequence of two-base repeat regions in tropical tree genomes. Genome 34:66-71

Danesh D, Aaron S, McGill GE, ND Young (1994) Genetic Dissection of oligogenic resistance to bacterial wilt in tomato. Molecular plant-microbe interactions 7:464-471

Davisson MT, PA Lalley, J Peters, DR Doolittle, AL Hillyard, AG Searle (1991) Report of the comparative committee for human, mouse and other rodent. Cytogenet. Cell Genet 58:1152-1189

Darling SM, CM Abbott (1992) BioEassys 14:359-366

Diers BW, L Mansur, J Imsande, RC Shoemaker (1992a) Mapping of phytophthora resistance loci in soybean with restriction fragment length polymorphism markers. Crop Sci 32:377-383

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

Dietrich W, H Katz, SE Lincoln, HS Shin, J Friedman, NC Dracopoli, ES Lander (1992) A genetic map of the mouse suitable for typing intraspecific crosses. Genetics 131:423-447

Donis-Keller H, P Green, C Helms, S Cartinhour, B Weiffenbach, K Stephen, TP Keith, DW Bowden, DR Smith, ES Lander, D Botstein, G Akots, DS Rediker, T Gravius, VA Brown, MB Rising, C Parkers, JA Powers, DE Watt, ER Kauffman, A Bricker, R Phipps, H Muller-kahle, TR Fulton, S Ng, JW Schumm, JC Braman, RG Knowlton, Barker DF, SM Crooks, SE Lincoln, MJ Daly, J Abrahamson (1987) A genetic linkage map of the human genome. Cell 51:319-337

Doyle JJ (1988) 5S ribosomal gene variation in the soybean and its progenitor. Theor Appl Genet 75:621-624

Dudley JW, MA Saghai Maroof, GK Rufener (1992) Molecular marker information and selection of parents in corn breeding programs. Crop Sci 32:301-304

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

Emerson RA, GW Beadle, AC Fraser (1935) A summary of linkage studies in maize. Cornell Univ. Agr. Exp. Sta. Memoir 180

Everson EH, CW Schaller (1955) The genetics of yield differrences associated with awn barbing in the barley hybrid (Lion X Atlas) X Atlas. Agron J 47:276-280

Fatokun CA, DI Menancio-Hautea, D Danesh, ND Young (1992) Evidence for orthologous seed weight genes in cowpea and mung bean based on RFLP mapping. Genetics 132:841-846

Fitch WM (1977) Phylogenies constrained by the crossover process as illustrated by human hemoglobins and athirteen-cycle, eleven-amino-acid repeat in human apolipo protein A-I. Genetics 86:623-644

Godziker T, J Williams, P Sharp, J Sambrook (1974) Physical mapping of temperature-sensitive mutations of adenoviruses. Cold Spring Harbor Symp Quant Biol 39:439-436

Gusella JF, NS Wexler, PM Conneally, SL Naylor, MA Anderson, RE Tanzi, PC Watkins, K Ottina, MR Wallace, AY Sakaguichi, AB Young, I Shoulson, E Bonilla, JB Martin (1983) A polymorphic DNA marker genetically linked to Huntington's desease. Nature 306:234-238

Hamada H, MG Petrino, T Kakunaga (1982) A novel repeat element with Z-DNA forming potential is widely found in evolutionarily diverse eukaryotic genomes. Proc Natl Acad Sci USA 79:6465-6469

Helentjaris T, G King, M Slocum, C Siedenstrang, S Wegman (1985) Restriction fragment length polymorphisms as probes for plant diversity and their development as tools for applied plant breeding. Plant Mol. Biol. 5:109-118

Helentjaris T (1987) A genetic linkage map for maize based on RFLPs. Trends in Genet 3:217-221

Jeffreys AJ (1979) DNA sequence variants in the G-gamma-A-gamma-, sigma- and beta-globin genes of man. Cell 18:1-10

Keim P, RC Shoemaker, RG Palmer (1989) RFLP diversity in soybean. Theor Appl Genet 77:786-792

Keim P, BW Diers, TC Olson, RC Shoemaker (1990a) RFLP Mapping in soybean: Association between marker loci and variation in quantitative traits. Genetics 126:735-742

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

Kiem P, W Beavis, J Schupp, R Freestone (1992) Evaluation of soybean RFLP marker diversity in adapted germ plasm. Theor Appl Genet 85:205-212

Kreike CM, JRA de Koning, JH Vinke, JW van Ooijen, C Gebhardt, WJ Stiekema (1993) Mapping of loci involvedin quantitatively inherited resistance to the potato cyst nematode *Globodera rostochiensis* pathotype1. Theor Appl Genet 87:464-470

Kurata N, G Moore, Y Nagamura, T Foote, M Yano, Y Minobe, M Gale (1994) Conservation of genome structure between rice and wheat. Bio/Tech 12:276-278

Landau-Ellis D, S Angermuller, RC Shoemaker, PM Gresshoff (1991) The genetic locus controlling supernodulation in soybean (*Glycine max* L.) co-segregates tightly with a cloned molecular marker. Mol Gen Genet 228:221-226

Lande R, R Thompson (1990) Efficiency of marker-assisted selection in the inprovement of quantitative traits. Genetics 124:743-756

Lander ES, D Botstein (1989) Mapping mendelian factors underlying quantiative traits using RFLP linkage maps. Genetics 121:185-199

Landry BS, RC Kesseli, B Rarrara, RW Michelmore (1987) A genetic map of lettuce (*Lactuca sativa* L.) with restriction fragment length polymorphism, isozyme, disease resistance and morphological markers. Genetics 116:331-337

Langercrantz U, H Ellegren, L Andersson (1993) The abundance of various polymorphic mirosatellite motifs differs between plants and vertebrates. Nucleic Acids Res. 21:1111-1115

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

Litt M, JA Luty (1989) A hypervariable microsatellite revealed by in vitro amplification of adinucleotide repeat within the cardiac muscle actin gene. Am J Hum Genet 4:397-401

Little P, G Annison, S Darling, R Williamson, T Cambar (1980) Model B: Model for antenatal diagnosis of β-thalassemia and other monogenic disorders by molecular analysis of linked DNA polymorphisms. Nature 285:144-147

MacArthur JW (1934) Linkage groups in tomato. J. Genet. 29123-133

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

Martin GB, SH Brommonschenkel, J Chenwongse, A Frary, MW Ganal, R spivey, T Wu, ED Earle, SD Tanksley (1993) map-based cloning of a protein kinase gene conferring disease resistance in tomato. Science 262:1432-1436

McCouch SR, G Dochert G, ZH Yu, ZY Zhang, GS Kush, WR Coffman, SD Tanksley (1988) Molecular mapping of rice chromosomes. Theor Appl Genet 76:815-829

Menancio-Hautea D, CA Fatokun, L Kumar, D Danesh, ND Young (1993) Comparative genome analysis of mungbean (*Vigna radiata* L. Wilczek) and cowpea (*V. unguiculata* L. Walpers) using RFLP mapping data. Theor Appl Genet 86:797-810

Mendel G (1865) Versuche über Pflanzen-Hybriden. Verh. Naturf.-Ver. Brunn IV:3-47

Morgante M, AM Olivieri (1993) PCR-amplified microsatellites as markers in plant genetics. Plant J 3:175-182

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

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

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

Phillips J, S Panny S, H Dazazian, C Bochun, C Scott, R Smith (1980) Prenatal diagnosis of sickle cell anemia by restriction endoclease analysis: HindIII polymorphism in v-globin gene extend applicability. Proc Nat Acad Sci USA 77:2853-2856

Rasmusson JM (1935) Studies on the inheritance of quantitative characters in *Pisum*: I. Preliminary note on the genetics of flowering. Hereditas 20:161-180

Saghai Maroof MA, RM Biyashev, GP Yang, Q Zhang, RW Allard (1994) Extraordinarily polymorphic microsatellite DNA in barley: Species diversity, chromosomal locations and population dynamics. Proc Natl Acad Sci USA 91:5466-5470

Sax K, (1923) The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics 8:552-560

Schön CC, AE Melchinger, J Boppenmaier, E Brunklaus-Jung, RG Herrmann, JF Seitzer (1994) RFLP mapping in maize: Quantitative trait loci affecting testcross performance of elite European flint lines. Crop Sci 34:378-389

Senior ML, Heun M (1993) Mapping maize microsatellites and polymerase chain reaction confirmation of the targeted repeats using a CT primer. Genome 36:884-889

Serikawa T, T Kuramoto, P Hilbert, M Mori, J Yamada, CJ Dubay, K Lindpainter, D Ganten, J Guénet, GM Lathrop, JS Beckmann (1992) Rat gene mapping using PCR-analyzed microsatellites. Genetics 131:701-721

Shoemaker RC, RD Guffy, LL Lorenzen, JE Specht (1992) Molecular genetic mapping of soybean: Map utilization. Crop Sci 32:1091-1098

Soller M, JS Beckmann (1983) Genetic polymorphism in varietal identification and genetic improvement. Theor Appl Genet 67:25-33

Starling RL, AF Ford, D Nelson, DC Torney, CD Hildebrand, RK Moyzis (1991) Evolution and distribution of (GT)n repetitive sequences in mammalian genomes. Genomics 10:807-815

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

Tanksley SD, TJ Orton (eds) (1983): Isozymes in plant genetics and breeding, Parts 1A and 1B. Elsevier, Amsterdam

Tautz D (1989) Hypervariablity of simple sequences as a general source of polymoprhic DNA markers. Nucleic Acids Res 17:6463-6471

Thoday JM (1961) Location of polygenes. Nature 191:368-370

Wang ZY, G Second, SD Tanksley (1992) Polymorphism and phylogenetic relationships among species in the genus *Oryza* as dtermined by analysis of nuclear RFLPs. Theor Appl Genet 83:565-581

Weber JL, PE May (1989) Abundant calss of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 44:388-396

Weber JL (1990) Informativeness of human (dC-dA)n (dG-dT)n polymorphisms. Genomics 7:524-530

Weeden NR, FJ Muehlbauer, G Ladizinsky (1992) Extensive conservation of linkage relationships between pea and lentil genetic maps. J Hered 83:123-129

Weissenbach J, G Gyapay, C Dib, A Virgnal, J Maoissette, P Millasseau, G Vaysseix, M Lathrop (1992) A second-generation linkage map of the human genome. Nature 359:794-801

Whitkus R, J Doebley, M Lee (1992) Comaprative genome mapping of sorghum and maize. Genetics 132:1119-1130

Wu KS, SD Tanksley (1993) Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol Gen Genet 241:225-235

Yang GP, Saghai Maroof MA, Xu CG, Zhang Q (1994) Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. Mol. Gen. Genet. 245:187-194

Young ND, D Zamir, MW Ganal, SD Tanksley (1988) Use of isogenic lines and simultaneous probing to identfy DNA markers tightly linked to the *Tm-2a* gene in tomato. Genetics 120:579-585

Yu YG, MA Saghai Maroof, GR Buss, PJ Maughan, SA Tolin (1994) RFLP and microsatellite mapping of a gene for soybean mosaic virus resistance. Phytopathology 84:60-64

Zhang Q, YJ Gao, SH Yang, R Ragab, MA Saghai Maroof, ZB Li (1994a) A diallel analysis of heterosis in elite hybrid rice based on RFLPs and microsatellites. Theor Appl Genet (in press)

Zhang Q, YJ Gao, MA Saghai Maroof, SH Yang, JX Li (1994b) Molecular divergence and hybid performance in rice. Plant Mol Biol (in press)

Zhao X, Kochert G (1992) Characterization and genetic mapping of a short, highly repeated, interspersed DNA sequence from rice (*Oryza sativa* L.). Mol Gen Genet 231:353-359

CHAPTER II.

Microsatellite and amplified sequence length polymorphisms in cultivated and wild soybean

ABSTRACT

The objectives of this study were to: i) assess the extent of genetic variation in soybean microsatellites (simple sequence repeats or SSRs), ii) assay for amplified sequence length polymorphisms (ASLPs), and iii) evaluate the usefulness of SSRs and ASLPs as genetic markers. Five microsatellites detected a total of 79 variants (alleles) in a sample of 94 accessions of wild (*Glycine soja*) and cultivated soybean (*G. max*). F_2 segregation analysis of four of the five microsatellites identified these variants (alleles) at four loci located in independent linkage groups. The number of alleles per microsatellite locus ranged from five to 21. To our knowledge, these are the highest number of alleles for a single Mendelian locus reported in soybean. Allelic diversity for the SSR loci was greater in wild than in cultivated soybean. Overall, 43 more SSR alleles were detected in wild than in cultivated soybean. These results indicated that SSRs are the marker of choice, especially for species with low levels of variation as detected by other types of markers. Two alleles were detected at each of the three ASLP loci examined. A total of six ASLP alleles were observed in cultivated soybean, and five were observed in wild soybean. All alleles detected in wild soybean were present in cultivated soybean. Allelic diversity values for the ASLP loci were near previous estimates for RFLPs.

Key words: Microsatellite - Simple sequence repeat - Soybean - Amplified sequence length polymorphism - Genetic mapping

INTRODUCTION

Microsatellites or simple sequence repeats (SSRs) consist of tandemly repeated core sequences that often vary in repeat number and are flanked by conserved DNA sequences. Thus, primers complementary to the flanking regions can be used to amplify the locus via the polymerase chain reaction (PCR) analysis. Variation in the microsatellite region resulting from variable number of repeated core sequences are then resolved electrophoretically on standard sequencing gels. SSRs are ideal genetic markers in that they are: i) highly abundant and evenly distributedin the genome, ii) highly polymorphic, iii) codominant, iv) rapidly typed via PCR, and v) very accessible to other laboratories via published primer sequences (Weber 1990; Saghai Maroof et al. 1994). While microsatellite variation has been increasingly exploited in mammalian systems, plant microsatellites remain virtually unexamined.

Condit and Hubbell (1990) assayed the abundance of dinucleotide repeats in maize and five tropical tree genomes. Akkaya et al. (1992) studied SSR variation in 43 soybean accessions and found that the number of alleles at three loci ranged from six to eight. Morgante and Olivieri (1993) surveyed published DNA sequence data for the presence and abundance of dinucleotide and trinucleotide repeats. Such SSRs were identified in 34 species with a frequency of one every 50 kilobases (kb). Wu and Tanksley (1993) observed from five to 11 alleles at eight microsatellite loci in a sample of 20 rice accessions. The Mendelian inheritance of four maize microsatellite loci was shown by Senior and Heun (1993). More recently, Saghai

Maroof et al. (1994) reported the inheritance, polymorphism, and population dynamics of barley microsatellite sequences and found that the number of alleles ranged from three to 37. Yu et al. (1994a) reported the first genetic mapping of a gene conferring disease resistance in plants using a soybean microsatellite marker, and Zhang et al. (1994) used microsatellites to identify chromosomal regions in rice that have significant effects on grain yield.

Amplified sequence length polymorphism (ASLP) is a technique to convert known DNA sequences to PCR-based genetic markers (Yu et al. 1994b). It was reasoned that PCR-amplified DNA fragments, which differ in sequence length due to random insertion-deletion mutations, may be an alternate source of DNA polymorphism. In short, primers are designed (without regard to internal sequence novelties) to amplify noncoding regions of DNA sequences using the polymerase chain reaction. Subsequent resolution of the resulting DNA fragments on standard sequencing gels allows for the detection of sequence length polymorphisms. Such amplified sequence length polymorphisms (ASLPs) are inherited in a Mendelian fashion and can be easily integrated into existing genetic maps. The utility of this approach was demonstrated by Yu et al. (1994b) in an effort to identify genetic markers flanking a resistance gene (*Rsv*1) for soybean mosaic virus. They analyzed seven potential ASLP loci in noncoding regions of several gene sequences residing in the vicinity of the virus resistance gene. Among the seven ASLP's used, two were closely linked to the resistance gene. In the present study we will assume ASLP

variation to be the result of random insertion-deletion mutations or transposition events, and simple sequence repeat variation as the result of mismatched and unequal crossing-over during meiosis resulting in varying numbers of tandem repeats at the SSR loci.

Here we report the extent of genetic variation for five microsatellite and three ASLP loci in cultivated soybean (*Glycine max*) and in its putative wild ancestor (*G. soja*). We also report the Mendelian inheritance and chromosomal locations of the microsatellite markers and discuss their utility as genetic markers.

MATERIALS AND METHODS

Genetic Material: Three sets of experimental soybean materials were used. The first set was composed of a broad sample of 62 *G. max* (hereafter Gm) accessions including ancestral cultivars, modern cultivars, breeding lines, and plant introductions, ranging from maturity group III to VIII with the majority of the accessions being from maturity groups IV and V. The second set included 32 *G. soja* (wild soybean, hereafter Gs) plant introductions from diverse origins in the Far East, including China, Japan, Korea, Taiwan, and the former Soviet Union. Cultivars and breeding lines were obtained from the soybean breeding program at Virginia Polytechnic Institute and State University, while plant introductions were kindly provided by Dr. R. Nelson, USDA-ARS, at the University of Illinois. The third set of genetic

materials included 150 F_2 individuals of an interspecific cross between a *G. max* breeding line (V71-370) and a *G. soja* plant introduction (PI 407.162).

Microsatellite and ASLP Sequences and Primers: A search of all available soybean sequences in GeneBank and EMBL databases identified five microsatellite-containing sequences, three of which (SOYPRP1, GmHSP176L and SOYSC514, hereafter the SOY or Gm prefixes will be deleted) have been described elsewhere (Akkaya et al. 1992; Yu et al. 1994a). The fourth and fifth loci are, HSP179D, a heat shock protein (GmHSP 17.9-D, class II) gene, with $(AT)_{10}$ (Raschke et al. 1988) and HSP173B, a heat shock protein (GmHSP 17.3-B) gene, with $(A)_{14}$ (Schoeffl et al. 1984). Primer sequences of these two loci are presented in Table 1.

The targeted sequences for ASLP amplification in this study were limited to noncoding regions of three heat shock protein genes. Two reasons for the use of heat shock protein genes are: i) they are known to belong to multigene families and, therefore, might be more variable than other gene sequences and ii) we were interested in developing additional markers linked to a gene conferring resistance to soybean mosaic virus that was previously reported to be linked to a heat shock protein multigene family (Yu et al. 1994a). Three heat shock protein genes were selected for detailed ASLP studies: i) HSP185C (GmHSP18.5-C, Raschke et al. 1988), ii) HSP175E (GmHSP17.5-E, Czarnecka et al. 1985), and iii) HSP175M (GmHSP17.5-M, Nagao et al, 1985). Primer pairs amplifying noncoding regions for

each of the three heat shock protein genes were designed using the computer program Primer 5.0 (Whitehead Institute, Cambridge, MA) and synthesized by Operon Technologies, Alameda, CA. The primer sequences of HSP185C and HSP175E have been described by Yu et al. (1994b), and that of HSP175M is presented in Table 1.

Microsatellite and ASLP Assays: Soybean DNA samples were prepared from freeze-dried tissue according to previously described procedures (Saghai Maroof et al. 1984). SSR assays were conducted according to Yu et al. (1994a) and Saghai Maroof et al. (1994). PCR products were denatured for 7 min at 94°C and separated on 7% denaturing polyacrylamide gels with 8 M urea at 1800 V constant voltage using a standard DNA sequencing unit (Model STS-45, IBI, Newhaven, CT). Gels were immediately covered with plastic wrap and exposed to X-ray film for 30-60 min. Procedures for ASLP analysis are essentially the same as those described for SSR analysis. Because ASLP insertion-deletion polymorphism can range from as few as nine bp to greater than 90 bp (see Results Section), the length of the gel running time was optimized for each locus to obtain maximum resolution (normally between 2-5 hrs).

Statistical Analysis: Diversity values for each locus were calculated using the genetic diversity index, $h=1-\Sigma p_i^2$, where p_i is the frequency of the ith allele. The diversity

values for the total sample were further partitioned into their respective components as described by Nei (1973), $H_T = H_W + G_{ST}$, where H_T is the total diversity, H_W is the within-group diversity component and G_{ST} is the between-group differentiation component.

RESULTS

Microsatellite polymorphism in the total sample. Analysis of the 62 Gm and 32 Gs DNA samples with the five microsatellite primer pairs identified a total of 79 SSR variants. Segregation of the banding patterns of the microsatellites (see Inheritance Section) conforms to a single Mendelian locus; therefore, we will refer to the target sequences of each primer pair as a locus and each variant as an allele.

Variation was detected at all five SSR loci in the total sample. An example of typical SSR variation at a single lucous in an arbituary sample of soybean accessions is shown in Figure 1. The number of alleles observed at each locus varied from a low of five for HSP173B to a high of 21 HSP176L (Table 2). PCR amplified fragments corresponding to different alleles of each SSR locus were run side by side on a standard sequencing gel. An allelic ladder (series) was constructed for each locus by resolving the observed alleles from smallest to largest (size in bp) on a sequencing gel. Identification of the alleles according to their position on the allelic ladder provided the best means for determining the correct genotype of each

accession and the correct number of alleles observed at each locus. Figures 2 to 4 present the allelic series from the total sample formed at four of these loci. Except for a few missing variants (steps), the allelic series for loci PRP1, SC514 (Figure 2 A&B), HSP176L (Figure 3), and HSP173B (figure not presented) form continuous ladders. Sixteen alleles were detected at locus PRP1 (Figure 2A). In the total sample, one of these 16 alleles (allele #2, Figure 2A, Table 3) was observed at a frequency of 0.44 and a second (allele #13, Figure 2A, Table 3) at a frequency of 0.13. The remaining 14 alleles were observed at frequencies less than 0.10 (Table 3). Nineteen alleles were detected at locus SC514 (Figure 2B). One of these 19 alleles (allele #5, Figure 2B, Table 3) was observed at a frequency of 0.39 and one (alleles #13, Figure 2B, Table 3) at a frequency of 0.12, whereas the other 17 alleles were observed in frequencies at or lower than 0.10 (Table 3). Twenty-one alleles were detected at locus HSP176L (Figure 3). Two of these 21 alleles (alleles #3 and #4) were observed in frequencies of 0.33 and 0.30 (Table 3), while the remaining 19 alleles were observed in frequencies less than 0.10. Of the five alleles detected at locus HSP173B, one was observed at a frequency of 0.62 and a second at a frequency of 0.25 (Table 3). The remaining three alleles were observed at frequencies less than 0.10.

SSR bands (variants) corresponding to the alleles detected at the locus HSP179D form a discontinuous ladder that can be divided into three smaller continuous ladders (Figure 4). Three alleles compose the lower ladder, one of which is detected at a

frequency of 0.42, whereas the remaining two alleles are observed in frequencies of 0.01. The intermediate ladder is composed of four alleles, all derived from Gs and observed at frequencies of 0.01. The upper ladder is composed of 11 alleles ranging in frequencies between 0.01 to 0.21 (Table 3).

Comparison of Microsatellite Polymorphism in Gm and Gs. Among the 79 alleles detected, 16 occurred in both subspecies, ten were observed only in Gm, and 53 were observed only in Gs. Thus, 43 more alleles were detected in Gs than in Gm (Table 2).

Diversity values for the total set ranged from 0.55 to 0.81, while the average diversity value of Gm was 0.55 and that of Gs was 0.87. A statistical test (Zhang and Allard 1986) indicated that the diversity values for Gs were significantly higher than those of Gm for all five SSR loci (Table 2A). Partitioning the diversity of the total sample into a within-group diversity component and a between-group differentiation component indicated that approximately seven to 19% of the total variation at these loci was due to between-group differentiation (G_{ST}%) (Table 2A).

Inheritance of the Microsatellite Loci. Linkage relationships for four of the five microsatellite loci were determined using F_2 segregation data from 150 plants for 86 molecular marker loci from a soybean interspecific cross (Maughan et al., unpublished). The parents (V71-370 and PI 407.162) differed in SSR alleles at only

four (PRP1, SC514, HSP176L, and HSP179D) of the five loci. Chi-square tests indicated that at these four loci the observed values did not deviate from the expected 1:2:1 segregation ratio (χ^2 values were 1.9, 1.3, 1.1, and 0.7 for PRP1, SC514, HSP176L, and HSP179D, respectively. Therefore, the banding pattern observed for each primer pair conforms to a single Mendelian locus with multiple alleles. Linkage analysis located HSP176L in the region between pK644 and pA186 which corresponds to linkage group F of the USDA-ISU soybean linkage map. In our mapping population, PRP1 was tightly linked to one of two RFLP loci detected by probe pA65 at a distance of less than 1 centiMorgan (cM). Since, probe pA65 did not detect two loci in the USDA-ISU soybean linkage map, the exact linkage group in which PRP1 resides cannot be determined with certainty. None of our 86 molecular markers showed linkage with the HSP179D or SC514 locus. Figure 5 represents an example of the segregation pattern for the SSR locus HSP179D.

ASLP Polymorphism in the Total Sample. Analysis of the 62 Gm and 32 Gs samples at three ASLP loci identified six variants (Table 2B). In our mapping population, all three ASLP loci were monomorphic and, hence, could not be tested for Mendelian inheritance. However the banding patterns for each primer pair was in agreement with that of single locus variation. Furthermore, two of the three loci (HSP185C and HSP175E) were shown to segregate in a Mendelian fashion in a previous study (Yu

et al. 1994b). Therefore, variants detected by each primer pair are allelic at a single locus.

Two alleles were detected at each of three loci HSP185C, HSP175E and HSP175M; the most frequent allele for these loci in the total sample was observed at frequencies of 0.71, 0.87, and 0.84, respectively (Table 3). ASLP polymorphism at each of these three loci is presented in Figure 6. The size of the insertion-deletion mutation at these loci varied substantially from approximately nine bp (HSP175E) to greater than 90 bp (HSP175M).

Comparison of ASLP Polymorphism in Gm and Gs. Among the six polymorphic ASLP alleles detected, five occurred in both subspecies and one was observed only in Gm (Table 2B). Diversity values for these loci in the total sample ranged from 0.23 to 0.46, while the average diversity value of Gm was 0.30 and that of Gs was 0.21 (Table 2B). Statistical analysis indicated that the diversity values for HSP185C were significantly higher in Gm than in Gs and that the opposite was true for HSP175M. At the third locus (HSP175E), the diversity value was about equal in wild and cultivated soybean. The between-group differentiation component (G_{ST}%) accounted for little of the total variation (0-7%) at loci HSP175E and HSP175M; however, for HSP185C, 30.5% of the total variation was due to between-group differentiation.

DISCUSSION

We analyzed microsatellite (or SSR) and ASLP variation in a sample of 62 cultivated and 32 wild soybean accessions. Microsatellite variation was clearly detected in the experimental materials studied; the number of alleles detected per locus was very high, ranging from 5 to 21. The average number of alleles (15.8) detected at a single SSR locus is similar to 17.7 reported in barley (Saghai Maroof et al. 1994) but higher than nine detected by Yang et al. (1994) in rice. The highly polymorphic nature of soybean microsatellites as observed in this study clearly contrasts with the lower levels of soybean RFLP polymorphism (Keim et al. 1992; Apuya et al. 1988). In addition, the number of alleles per locus is much larger than those reported previously from studies in soybean using other types of genetic markers including allozymes (Perry et al. 1991), ribosomal DNA (Doyle, 1988), chloroplast DNA (Shoemaker et al. 1986), and mitochondrial DNA (Grabau et al. 1992).

Segregation analysis for four of the five SSR loci of this study indicated that they map to four separate linkage groups. SSR bands corresponding to alleles at each of four loci produced continuous allelic ladders (Figures 2 and 3), similar to those described by Saghai Maroof et al. (1994) in barley and Yang et al. (1994) in rice. This type of polymorphism is characteristic of genetic changes attributable to simple sequence repeat variation produced by mismatch and unequal crossing-over during meiosis. One of these loci (HSP173B) detected five alleles, each differing by one

base pair. The target sequence amplified at this locus contains a mononucleotide poly-A core motif $[(A)_{14}]$. The genetic diversity values for this locus are lower than those of the other four (di- and tri-nucleotide repeat) SSR loci (Table 2), and may indicate a lower level of genetic diversity for mononucleotide SSR loci. However, it should be noted that if mononucleotide variation is to be examined, amplified target sequences should be between 50 to 125 bp, due to the difficulty in resolving single base pair variation in larger DNA fragments.

SSR alleles at the fifth microsatellite locus (HSP179D) formed a novel discontinuous ladder (Figure 4). Of the 18 alleles observed at this locus, three were observed in the lower ladder, four in the intermediate ladder, and 11 in the upper ladder. Interestingly, none of the alleles of the intermediate ladder were derived from Gm. In addition, there are more than 100 bp separating the largest allele (#3) of the lower ladder and the smallest allele (#8) of the upper ladder (see Figure 4). There are several possible explanations for these observations, including: i) limited diversity for this locus in the soybean accessions studied-i.e., the missing alleles are present in soybean but not in the material studied, ii) there is a selective disadvantage for soybeans having the intermediate allelic types, and iii) the discontinuous distribution is the result of an insertion-deletion mutation at a location adjacent to the simple sequence repeat, reminiscent of ASLP variation.

Our SSR survey clearly showed that wild soybean (Gs) is more variable than cultivated soybean (Gm), both in number of alleles and in genetic diversity values.

In the total sample, 43 more alleles were detected in Gs than in Gm. Interestingly, many alleles observed at high frequencies in Gm are not observed or observed only in low frequencies in Gs. For example, the most frequent allele (#5) in Gm for the locus SC514 (*f*=0.59) is not observed in our sample of Gs (Table 3). Changes in the number of nucleotide repeat units at an SSR locus may have direct effects on the gene in which they reside and therefore may be subject to artificial or natural selection. Indeed, researchers have shown a direct correlation between the number of SSR repeats and severity in six human genetic diseases (Wang et al. 1994). Dramatic changes in SSR allelic frequencies also were observed by Saghai Maroof et al. (1994) who examined two generations (F_8 and F_{53}) of an experimental barley population. Such changes were explained on the basis of natural selection acting on the chromosomal segments in which the marker loci resided.

In addition to surveying microsatellite variation, we also observed allelic frequencies of three amplified sequence length polymorphisms (ASLPs). ASLP variation was clearly detected at all three loci surveyed. Two alleles were observed in both Gm and Gs for HSP175E and HSP175M, similar to the situation observed by Keim et al. (1992) for RFLP diversity in soybean, who observed diallelism for 88 RFLP markers among 91 genomic DNA probes. At the third locus (HSP185C), two equally abundant alleles were observed in Gm while only one allele was observed in Gs (Table 3). The size of the insertion-deletion mutation for each of these ASLP loci varied dramatically from approximately nine bp to greater than 90 bp. Whether

the source of ASLP variation results from random insertion-deletion mutations or from transposition events is unknown. Genetic diversity levels for these three loci are within the range of previous estimates for RFLPs (0.10-0.70) from the same set of soybean accessions (Maughan et al., unpublished data) and, therefore, may be useful as genetic markers in site directed mapping projects. The utility of ASLP markers has already been demonstrated by Yu et al. (1994b), who developed ASLP markers closely linked to a virus resistance gene in soybean.

In cultivated soybean the level of RFLP polymorphism is generally low and is one of the limiting factors hindering the development of high density genetic maps and their subsequent use in marker-assisted selection programs. The development of alternative methodologies for molecular marker development is essential for future progress in soybean genetics. ASLPs are unique in that they detect minor length changes in DNA fragments and are not restricted to DNA sequences containing repeated sequences. However, they are significantly less polymorphic than microsatellite markers and are restricted to DNA segments of known sequence. In contrast, the ease of use, highly polymorphic nature, and abundance of microsatellite markers make them the genetic marker of choice. Once developed they are easily maintained and shared among laboratories by publishing primer pairs. Thus, it is our expectation that SSRs will provide a rich new source of genetic polymorphism for all plant species, including soybean.

Acknowledgements. We thank Dr. KS Wu for his technical advice on microsatellite analysis. This research was supported in part by grants from the Virginia Center for Innovative Technology, Montague Farms, Inc., Virginia Soybean Board, and the US Department of Agriculture grant 94-37300-0364.

REFERENCES

Akkaya MS, Bhagwat AA, Cregan PB (1992) Length polymorphism of simple sequence repeat DNA in soybean. Genetics 132:1132-1139

Apuya NR, Frazier BL, Keim P, Jill E, Lark KE (1988) Restriction fragment polymorphisms as genetic markers in soybean, *Glycine max* (L.) Merr. Theor Appl Genet 75:889-901

Czarnecka E, Gurley WB, Nagao RT, Mosquera LA, Key JL (1985) DNA sequence and transcript mapping of a soybean gene encoding a small heat shock protein. Proc Natl Acad Sci USA 82:3726-3730.

Condit R, Hubbell SP (1991) Abundance and DNA sequence of two-base repeat regions in tropical tree genomes. Genome 34:66-71

Doyle JJ (1988) 5S ribosomal gene variation in the soybean and its progenitor. Theor. Appl. Genet. 75:621-624

Grabau EA, Davis WH, Phelps ND, Gengenbach BG (1992) Classification of soybean cultivars based on mitochondrial DNA restriction fragment length polymorphisms. Crop Sci. 32:271-274

Keim P, Beavis W, Schupp J, Freestone R (1992) Evaluation of soybean RFLP marker diversity in adapted germplasm. Theor Appl Genet 85:205-212

Morgante M, Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics. The plant J. 3:175-182

Nagao RT, Czarnecka E, Gurley WB, Schoeffl F, Key JL (1985) Genes for low-molecular-weight heat shock proteins of soybeans: Sequence analysis of a multigene family. Mol. Cell. Biol. 5:3417-3428

Raschke E, Baumann G, Schoeffl F (1988) Nucleotide sequence analysis of soybean small heat shock protein genes belonging to two different multigene families. J. Mol. Biol. 199:549-557

Saghai Maroof MA, Soliman KM, Jorgenson RA, Allard RW (1984) Ribosomal DNA spacer length polymorphism in barley: Mendelian inheritance, chromosomal location and population dynamics. Proc Natl Acad Sci USA 81:8014-8018

Saghai Maroof MA, Biyashev RB, Yang GP, Zhang Q, Allard RW (1994) Extraordinarily polymorphic microsatellite DNA in barley: species diversity, chromosomal locations, and population dynamics. Proc Natl Acad Sci USA 91:5466-5470

Schoeffl F, Raschke E, and Nagao RT (1984) The DNA sequence analysis of soybean heat-shock genes and identification of possible regulatory promoter elements. EMBO J. 3:2491-2497

Senior ML, Heun M (1993) Mapping maize microsatellites and polymerase chain reaction confirmation of the targeted repeats using a CT primer. Genome 36:884-889

Shoemaker RC, Hatfield PM, Palmer RG, Atherly AG (1986) Chloroplast DNA variation in the genus *Glycine* subgenus *Soja.* J. Hered. 77:26-30

Wang YH, Amirhaeri S, Kang S, Wells RD, Griffith JD (1994) Preferential nucleosome assembly at DNA triplet repeats from the myotonic dystrophy gene. Science 265:669-671

Weber JL (1990) Informativeness of human $(dC\text{-}dA)_n$ $(dG\text{-}dT)_n$ polymorphisms. Genomics 7:524-530

Wu KS, Tanksley SD (1993) Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol. Gen. Genet. 241:225-235

Yang GP, Saghai Maroof MA, Xu CG, Zhang Q (1994) Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. Mol. Gen. Genet. 245:187-194

Yu YG, Saghai Maroof MA, Buss GR, Maughan PJ, Tolin SA (1994a) RFLP and microsatellite mapping of a gene for soybean mosaic virus resistance. Phytopathology 84:60-64

Yu YG, Saghai Maroof MA, Buss GR, Maughan PJ, Tolin SA, Ma G (1994b) A soybean mosaic potyvirus resistance gene is closely linked to a multigene family. (Submitted)

Zhang Q, Allard RW (1986) Sampling variance of genetic diversity index. J. Heted. 77:54-55

Zhang Q, Gao YJ, Yang SH, Ragab R, Saghai Maroof MA, Li ZB (1994) A diallel analysis of heterosis in elite hybrid rice based on RFLPs and microsatellites. Theor. Appl. Genet. (in press)

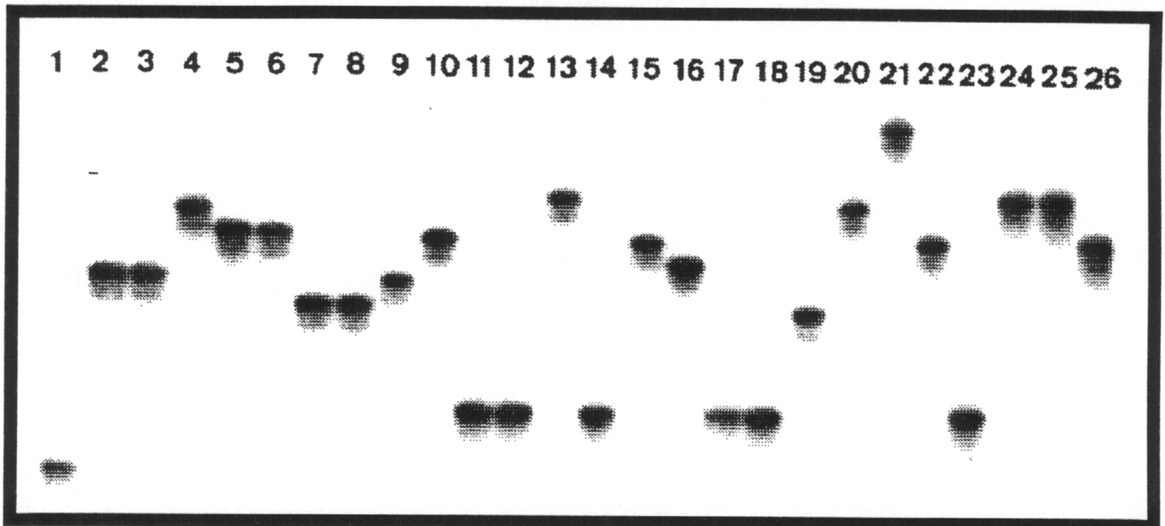


Figure 1. Microsatellite polymorphism at locus PRP1. Genomic DNA from 26 arbitary individuals was amplified via the polymerase chain reaction in the presence of ^{32}P-dCTP. PCR fragments were resolved on a standard DNA sequencing gel and detected by autoradiography.

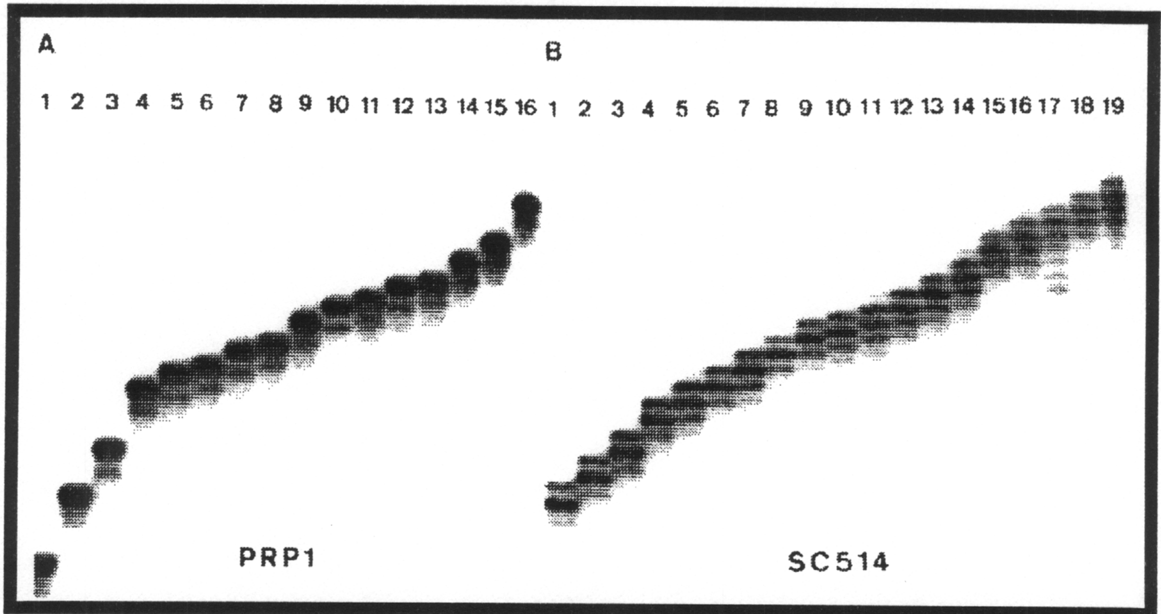


Figure 2. Continuous allelic ladders for two soybean microsatellite loci (A) PRP1 and (B) SC514. Genomic DNA from a single soybean accession representing each allele in the allelic series was amplified via the polymerase chain reaction in the presence of ^{32}P-dCTP. The allelic ladder was constructed by resolving the alleles from smallest to largest on a standard sequencing gel. Sixteen and 19 alleles were observed for microsatellite loci PRP1 and SC514, respectively. Allelic designations, as referenced in the text, are as numbered in the figure.

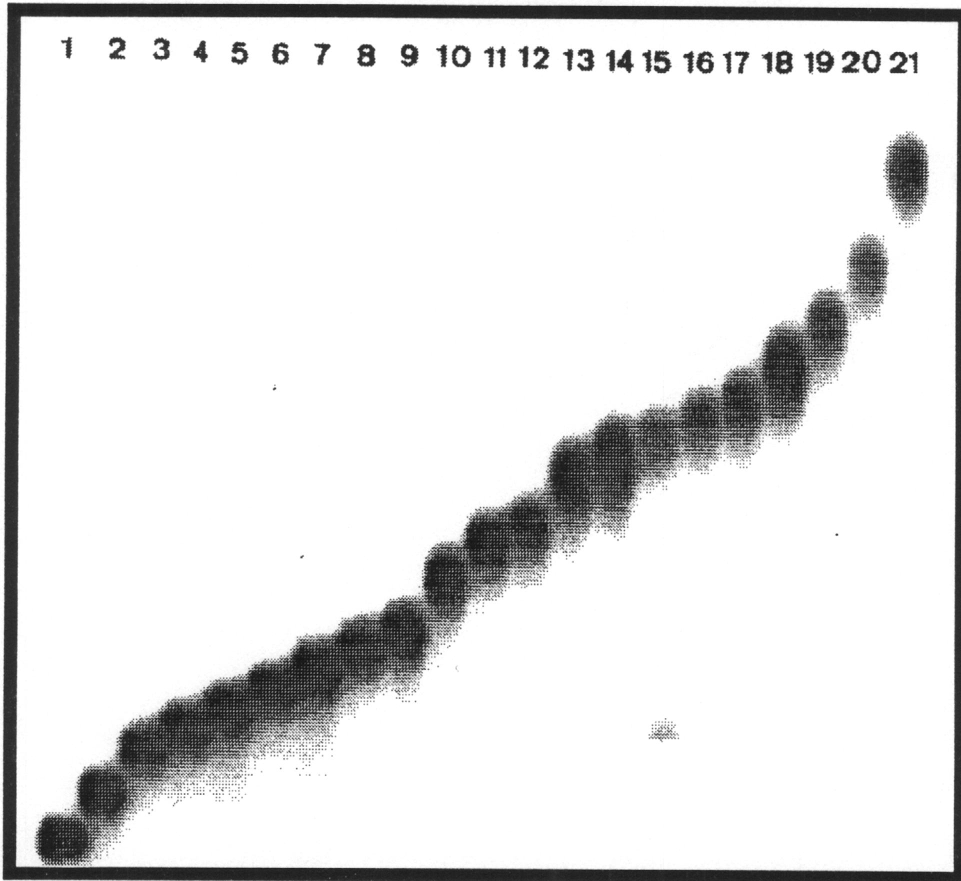


Figure 3. Continuous allelic ladder for the soybean microsatellite locus HSP176L. Twenty-one alleles observed at this locus were amplified and resolved as described in Figure 2. Allelic designations are as specified on top of the figure.

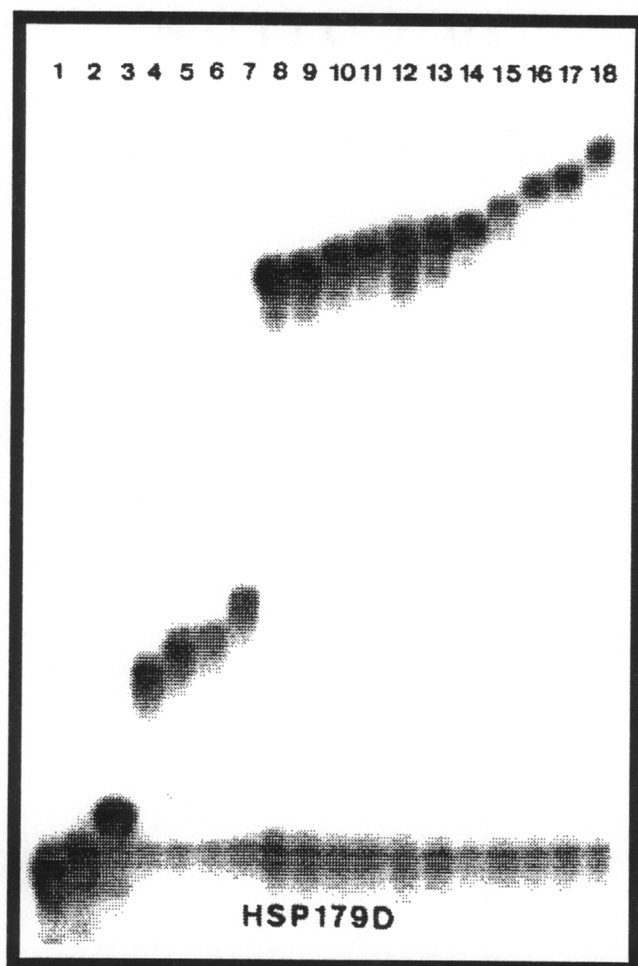


Figure 4. Discontinuous allelic ladder for the soybean microsatellite HSP179D. Genomic DNA from a single soybean accession representing each allele in the allelic series was amplified and resolved as described in Figure 2. The eighteen alleles detected at this locus produce three distinct and separate continuous ladders. Alleles 1-3, 4-7, and 8-18 create the lower, intermediate and higher ladders, respectively. Alleles are designated as numbered in the figure.

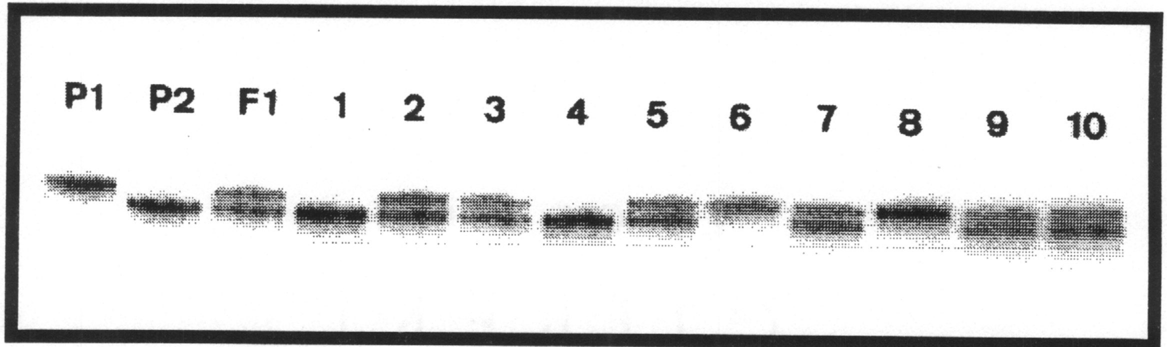


Figure 5. F_2 segregation pattern for the microsatellite HSP179D. Genomic DNA from the parental lines, the F_1 hybrid, and 10 random F_2 individuals were amplified using primers flanking an internal simple sequence repeat $[(TA)_{10}]$ in the presence of ^{32}P-dCTP. Amplified fragments were resolved on standard sequencing gels and detected by autoradiography. The first two lanes contain parental DNA, PI 407.162 and V71-370, respectively; the third lane contains DNA from the F1 hybrid, the remaining 10 lanes are a portion of the F_2 population constructed from this cross. The allelic difference between the parents is 2 bp.

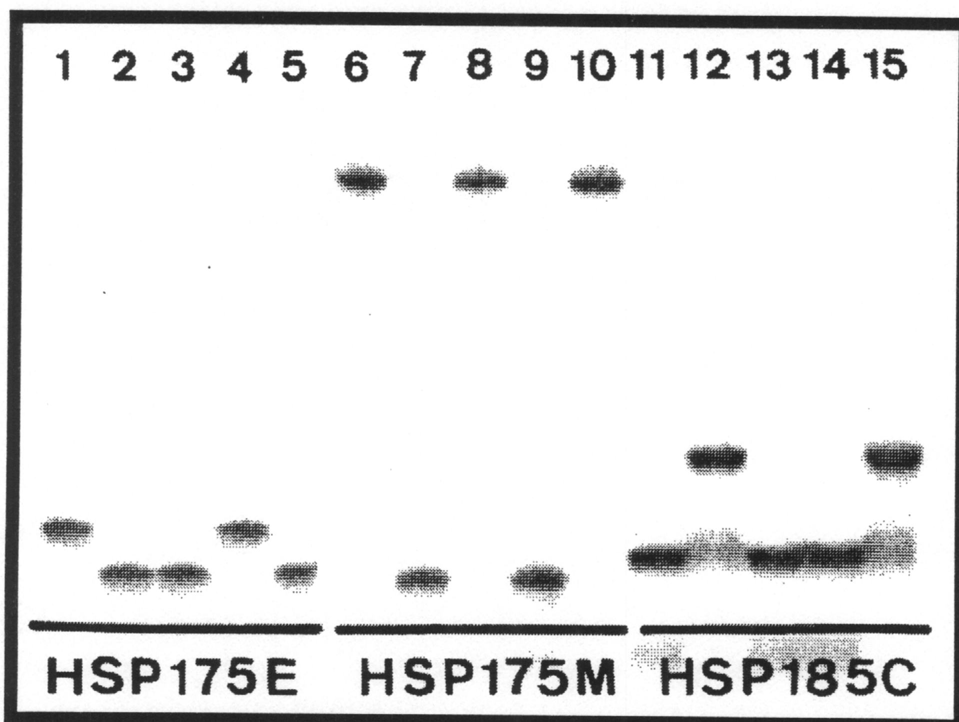


Figure 6. DNA polymorphism at three ASLP loc in Gm and Gs. Genomic DNA from five arbitarily selected individuals was amplified in the presence of ^{32}P-dCTP for each ASLP locus. PCR fragments were resolved on a standard DNA sequencing gel and detected by autoradiography. Lanes 1-5, 6-10, and 11-15, represent the genetic variation detected at ASLP loci HSP175E, HSP175M, and HSP185C, respectively.

Table 1. Oligonucleotide primer sequences designed to detect SSR[a] and ASLP variation in soybean.

Locus	Primer sequence (5′-3′)	Product length (bp)	Variant class
HSP173B	ATTGAGAAATCTAGATAGTCAGCCTT GACATTGGTTAGGATTGGTCTACC	227	SSR
HSP179D	TTGTTTCATTACGTGACTTCATCTA GGTTTGAATTGATTTTAAAACATCG	189	SSR
HSP175M	TTTTTTCCACTTCAGTTCCTCC TCTTGTTCTAGTTCTGGTTGATATCG	274	ASLP

[a]Simple sequence repeat core motifs for the microsatellites HSP173B and HSP179D are $(A)_{14}$ and $(TA)_{10}$, respectively.

Table 2. Number of alleles, levels of diversity, and differentiation in *Glycine max* and *Glycine soja* detected at four SSR and five ASLP loci.

A. SSR Locus	No. of alleles					Diversity values			
	Gm[a]	Gs	Total	Shared		Gm	Gs	Total	G_{ST}%[c]
PRP1	6	13	16	3		0.62	0.88*	0.76	7.1
SC514	8	16	19	5		0.62	0.90*	0.81	11.9
HSP176L	5	20	21	4		0.61	0.94*	0.79	8.7
HSP173B	2	5	5	2		0.29	0.75*	0.55	18.8
HSP179D	5	15	18	2		0.59	0.89*	0.7	9.0
Total	26	69	79	16	Avg	0.55	0.87	0.73	9.2
B. ASLP Locus									
HSP185C	2	1	2	1		0.49	0.00*	0.46	30.5
HSP175E	2	2	2	2		0.23	0.24	0.23	0.0
HSP175M	2	2	2	2		0.18	0.40*	0.27	6.8
Total	6	5	6	5	Avg	0.30	0.21	0.32	12.4
N[b]	62	32	94						

*Indicates significant difference between Gm and Gs

[a]Gm=*Glycine max*; Gs=*Glycine soja*

[b]Number of Gm and Gs individuals assayed

[c]G_{ST}=the between-group differentiation component

Table 3. Frequencies of the two most prevalent alleles of each SSR and ASLP locus in *Glycine max*, *Glycine soja* and the total sample.

Locus	Class	*G. max* Allele[a]	*G. max* Freq.	*G. soja* Allele	*G. soja* Freq.	Total Allele	Total Freq.
PRP1	SSR	2[a]	0.58	2	0.19	2	0.44
		10	0.14	9	0.19	13	0.13
		N[b]	6		13		16
SC514	SSR	5	0.59	2	0.23	5	0.39
		13	0.15	6	0.13	13	0.12
		N	8		16		19
HSP176L	SSR	3	0.47	2	0.13	3	0.33
		4	0.40	4	0.09	4	0.30
		N	5		20		21
HSP173B	SSR	1	0.82	1	0.37	1	0.62
		2	0.18	2	0.22	2	0.25
		N	2		5		5
HSP179D	SSR	2	0.55	2	0.19	2	0.42
		9	0.31	12	0.19	9	0.21
		N	5		15		18
HSP185C	ASLP	1	0.56	1	1.0	1	0.71
		2	0.46	2	-	2	0.29
		N	2		1		2
HSP175E	ASLP	1	0.87	1	0.89	1	0.87
		2	0.13	2	0.11	2	0.13
		N	2		2		2
HSP175M	ASLP	1	0.90	1	0.72	1	0.84
		2	0.10	2	0.28	2	0.16
		N	2		2		2

[a]SSR allele numbers are as designated in Figures 1-3
[b]N=total number of alleles detected in each group

CHAPTER III.

Molecular Marker Analysis of Seed Weight: Genomic locations, gene action and evidence for orthologous evolution among three legume species

ABSTRACT

The objectives of this study were to use molecular and morphological markers to i) identify quantitative trait loci (QTL) controlling seed weight in soybean, ii) characterize the genetic basis of seed weight expression, and iii) determine whether soybean shares orthologous seed weight genes with cowpea and/or mung bean. An F_2 population was developed between a large-seeded *Glycine max* breeding line (V71-370) and a small-seeded *G. soja* plant introduction (PI407.162). DNA samples from 150 F_2 individuals were analyzed with 91 genetic markers, including RFLPs, RAPDs, and SSRs. Seed weight was analyzed in two generations (F_2 and $F_{2:3}$) by randomly sampling 100 seeds from each of 150 F_2 individuals and 150 $F_{2:3}$ lines from a replicated field trial. Markers associated with seed weight were identified using the computer program MapMaker-QTL and one-way analysis of variance. Three and five markers were significantly associated with seed weight variation ($P<0.01$) in the F_2 and $F_{2:3}$ generations, respectively. Tests for digenic epistasis revealed three significant epistatic interactions in both generations. In a combined analysis, these markers and epistatic interactions explained 50 and 60% of the phenotypic variation for seed weight in these generations. Comparison of our results in soybean (*Glycine*) with those reported in cowpea and mung bean (*Vigna*) indicated that both genera share orthologous seed weight genes. In both genera a genomic region, significantly associated with seed weight, spanned the same RFLP markers in the same linkage order. Moreover, a significant epistatic interaction between these marker loci (or

nearby QTL) and RFLP markers at unlinked seed weight QTLs was conserved in all three species. These results suggest that the exploitation of "comparative QTL mapping" is an invaluable tool for quantitative geneticists working with depauperate or poorly characterized plant systems.

Key words: Genetic mapping - Restriction fragment length polymorphism - Soybean -Seed Weight - Orthologous evolution

INTRODUCTION

In soybean (*Glycine max*) and other plant species, the majority of economically important agronomic characteristics, such as yield and most of its component traits, are controlled in a quantitative fashion. Until recently, plant breeders have relied on complex selection methods to improve specific quantitative traits. Such methods are expensive, time consuming and labor intensive. Recent advances in molecular genetics, in particular the advent of restriction fragment length polymorphism (RFLP) technology, have made possible the genetic dissection of many of these recalcitrant agronomic characteristics (Stuber 1992). RFLP technology has been successfully used in the study of a wide array of quantitative traits, including fruit characteristics in tomato (Patterson et al. 1991), heterosis in maize (Stuber et al. 1992), and quantitative disease resistance in potato (Kreike et al. 1993) and soybean (Concibido et al. 1994).

Seed weight, measured as mass per seed, is an important yield component in soybean (Burton 1987). Soybean seed weight is polygenically controlled and ranges from seven to 55 grams per 100 seeds. Soybean cultivars with either very small (<8 g/100 seeds) or very large seed sizes (>20 g/100 seeds) are used directly in the production of many oriental specialty food items, including tofu, natto, and miso. The demand for these "food" quality soybeans in the global market is steadily increasing at a rate of 3 to 5 percent per annum (Griffis and Wiedermann 1990). Soybean cultivars with these desired seed sizes are limited in number and generally

unadapted to U.S. growing areas. The increasing interest in adapted soybean cultivars that fit into this specific market has made breeding for seed weight a major objective of some breeding programs. Plant breeders have been successful in manipulating this trait, but the underlying genetic basis for seed weight inheritance has not been elucidated. Information on the association between genetic markers and seed weight should help breeders construct beneficial allelic combinations and accelerate the development of specialty cultivars.

RFLP mapping has previously been used to analyze seed weight in a variety of crop species. Most recently, the inheritance of kernel weight in two maize-teosinte hybrid populations was investigated by Doebley et al. (1994). In the two populations studied, they detected multiple quantitative trait loci (QTL) explaining approximately 70% of the phenotypic variance for kernel weight. In soybean, Mansur et al. (1993) utilized 154 genetic markers and detected a single unlinked QTL explaining 13% of the phenotypic variation for seed weight. Fatokun et al. (1992) identified two and four QTLs explaining 53 and 50% of the variation for seed weight in cowpea and mung bean, respectively. Interestingly, in both cowpea and mung bean, the genomic region with the greatest effect on seed weight spanned the same RFLP markers in the same linkage order, suggesting that this genomic region has been conserved through the evolution of these two species. Although soybean belongs to a different genus than cowpea and mung bean (*Vigna*), all three species are from the tribe

Phaseoleae of the family Leguminosae (Smartt 1990). Thus it is possible that the three species share orthologous seed weight genes (Tanksley et al. 1988).

The purpose of this research was to use genetic markers to i) identify quantitative trait loci for seed weight in soybean, ii) characterize the genetic basis of seed weight expression in soybean, and iii) to determine whether soybean shares orthologous seed weight QTLs with cowpea and/or mung bean.

MATERIALS AND METHODS

Genetic Materials. An F_2 population with a high level of genetic diversity for seed weight was developed from an interspecific hybridization between an adapted *G. max* breeding line (V71-370) with large seeds (24 g/100 seeds) and a *G. soja* plant introduction (PI407.162) with extremely small seeds (1.8 g/100 seeds). F_1 and F_2 plants were grown under normal greenhouse conditions at Virginia Polytechnic Institute and State University. Sixty F_3 seeds from each F_2 individual were scarified and planted in a randomized complete block design with two replications at the Virginia Crop Improvement Association's seed farm near Mt. Holly, Va. $F_{2:3}$ lines were planted in plots four ft long with 7.5 ft between rows. Since plants within $F_{2:3}$ lines were expected to segregate for maturity and shattering was expected, $F_{2:3}$ lines were harvested in bulk and dried in large open grain bags when 90% of the plants in the plot reached maturity. After drying, each plot ($F_{2:3}$ line) was threshed

separately to produce F_4 seed. F_2 genotypes for three morphological loci (T/t-pubesence color, L1/l1-pod color, and Pa1/pa1-pubesence type) were classified in the $F_{2:3}$ generation.

Molecular Marker Assay. DNA was isolated from leaf tissue of parental, F_1 , and 150 F_2 plants as previously described (Saghai Maroof et al. 1984). RFLP procedures were as described by Zhang et al. (1993). Briefly, eight μg of plant DNA were digested with each of five restriction enzymes (*BamH*I, *Dra*I, *EcoR*I, *EcoR*V, and *Hind*III) and electrophoresed on 0.8% agarose gels, followed by standard DNA transfer to nylon membranes via Southern blotting. Southern blots were hybridized overnight with randomly primed ^{32}P-dCTP labelled probes. Following hybridization, blots were washed four times, twice for 5 min at room temperature with 1 X SSC and 0.1% SDS and twice for 10 min at 65 °C with 0.1 X SSC and 0.1% SDS. After washing, blots were immediately covered with plastic wrap and exposed to X-ray film for three to five days.

DNA Clones. Two sets of DNA clones derived from three separate *Pst*I genomic DNA libraries were used in this experiment. The first set, consisting of 170 soybean clones, was selected based on the position of the clones in the current USDA-ISU soybean genetic map to provide maximum genome coverage (Diers et al. 1992). The second set or "comparative QTL mapping" set was selected based on previous reports placing them in the vicinity of putative QTLs controlling seed weight in cowpea

and/or mung bean (Fatokun et al. 1992). This set included 11 soybean clones, five mung bean clones, and one cowpea clone. DNA clones were generously provided by R.C. Shoemaker (Iowa State University, Ames) and N.D. Young (University of Minnesota, St. Paul, MN). A total of 84 clones was polymorphic between the parents of this cross. Clones derived from soybean, cowpea and mung bean libraries will be designated with the prefix sg, cg, or mg, respectively.

In addition to screening for RFLPs, a set of forty randomly amplified polymorphic DNA (RAPD) primers (Operon kits B and O; Alameda, CA) and eight simple sequence repeat (SSR or microsatellite) primers were screened for polymorphism according to the procedures outlined by Williams et al. (1990) and Yu et al. (1994), respectively. Of the forty RAPD primers screened, three produced reproducible polymorphic markers. RAPD markers will be given the prefix R, followed by the Operon kit designation (B or O), followed by the primer number. Of the eight SSR markers screened, four have been described by Akkaya et al. (1992) and four were described by Yu et al. (1994). Four SSR markers were polymorphic between the parents of this cross. SSR markers will be given the prefix sm.

Data Analysis and QTL mapping. Seed weights for two generations (F_2 and $F_{2:3}$) were determined by randomly sampling 100 seeds from each F_2 plant and each $F_{2:3}$ plot. Trait means, normality, parent-offspring correlations, and analysis of variance F-tests were determined using the computer program SAS (Statistical Analysis

Systems, Cary, NC). Heritability in standard units was calculated using the correlation coefficient, adjusted for inbreeding (Frey and Horner 1957; Smith and Kinman 1965). Segregation ratios for each marker in the F_2 population were tested for goodness of fit to a 1:2:1 or 3:1 ratio using the computer program Linkage-1 (Suiter et al. 1983). Most probable order and map distances were determined by multiple linkage analysis using the computer program MapMaker 2.0 at LOD=2.4 (logarithm to the base 10 of the likelihood odds ratio) and with a maximum Haldane distance of 50 centiMorgans (Landers et al. 1987). Molecular marker-QTL associations were analyzed using the computer program MapMaker-QTL 1.1b (Landers and Botstein 1989) and analysis of variance F-tests (see Keim et al. 1990). Significant LOD values (>2.0) and/or F-values (P<0.01) were interpreted to indicate cosegregation of putative QTLs for seed weight and genetic markers. Two-way analysis of variance F-tests were used to test for genetic interactions between marker loci. Significant marker loci and interactions were then combined in a multivariate linear regression model to determine their combined effect.

Since multiple one-way ANOVAs were performed on the same data set, there is a potential risk of committing a type I error (P=0.60). Lowering the significance level, however, creates the possibility of missing a weak linkage to an important QTL. To overcome this dilemma some researchers (Nienhuis et al. 1987) have suggested checking significant associations in a second set of progeny from the same

population. In this study we attempt to confirm our results by using two data sets generated from two generations of the same population (F_2 and $F_{2:3}$).

RESULTS

Phenotypic Analysis. Analysis of seed weight distribution in the F_2 and $F_{2:3}$ generations shows continuous variation (Figure 1). The lack of discrete classes is consistent with a quantitatively controlled trait. Seed weight ranged from 3.6 to 10.4 g/100 seeds in the F_2 generation and from 4.6 to 10.0 g/100 seeds in the $F_{2:3}$ generation. Seed weights equal to that of the parents were not observed in either generation. The mean of the F_2 and $F_{2:3}$ generations were 6.7 and 7.1 g/100 seeds, respectively. The expected midparental values for the F_2 and $F_{2:3}$ generations were 12.8 and 13.1 g/100 seeds, respectively. In both generations the population means were closer to the weight of the small-seeded parent, suggesting that in this population, small seed weight exhibits partial dominance. Tests for normality, as required for MapMaker-QTL and ANOVA, indicated that seed weight was normally distributed in both generations. Standard unit heritability, calculated from the parent-offspring correlation and adjusted for inbreeding, was 0.54 and is in the range of previous heritability estimates for seed weight (Burton, 1987). Figure 2 presents the regression of the $F_{2:3}$ generation onto the F_2 generation. The regression coefficient (r) was 0.81. To test for the effects of heterozygosity on seed weight,

percent heterozygosity for all 91 marker loci was regressed against F_2 plant and $F_{2:3}$ family seed weight (Figure 3). Heterozygosity did not significantly predict seed weight in either generation ($r^2<0.05$ in both generations).

QTL mapping. The low resolution map produced from the genetic data consisted of 77 markers, forming 21 linkage groups spanning 780 centiMorgans (Figure not presented). An additional 14 markers were not linked. Deviation from the expected 1:2:1 genotypic ratios were significant for three (3.3%) of the 91 markers scored (sgA487, P=0.02; sgK227, P=0.01; sgA955, P=0.003). These three loci were located on separate linkage groups, had excess *G. max* alleles, and were not significantly associated with seed weight variation in soybean.

Searches for seed weight QTLs in the F_2 generation with both MapMaker-QTL and one-way ANOVA identified the same three marker loci (Table 1). Individually these markers, sgA118, sgA816, and sgK385, explained 21, 7.8, and 10.5 of the total variation for seed weight. Two-way ANOVA tests for digenic interactions revealed three significant epistatic interaction terms (Table 2). In a combined analysis, the significant marker loci and interaction terms explained 50% of variation for seed weight.

In the $F_{2:3}$ generation, five genomic regions associated with seed weight were identified by both MapMaker-QTL and one-way ANOVA (Table 1). Individually, these markers explained from 7.1 to 14.2% of the total variation for seed weight

(Table 1). Of the five significant marker loci identified in the $F_{2:3}$ generation, two were detected in the F_2 generation at $P<0.01$ (markers sgA118 and sgA816) and three at $P<0.05$. Tests for epistatic interaction between the significant marker loci detected in the $F_{2:3}$ generation and all other non-linked loci identified three significant epistatic interaction. Two of these were also detected in the F_2 generation (sgA816*mgM185b and sgA118*sgA486). In a combined analysis, significant marker loci and interaction terms explained 60% of the total phenotypic variation for seed weight. Significant marker loci idenified in this study are located on five separate linkage groups and appear to be randomly scattered throughout the soybean genome (Table 1).

In both generations, all *G. max* alleles at significant loci were associated with greater seed weight. These results were expected since the *G. max* parent had significantly larger seeds than the *G. soja* parent. Gene action at individual marker loci was evaluated by comparing the fit of individual QTLs to three Mendelian models using MapMaker-QTL 1.1b (ie. either dominant, recessive, or additive). A one LOD reduction in likelihood indicates that a given type of gene action is unlikely. The modes of inheritance for the individual QTLs, presented in Table 1, could not be rejected as unlikely. Gene action of all six loci affecting seed weight was consistent with an additive model (Table 1). Three of the six QTLs (sgA118, sgA023, and sgK384) also conformed to a dominance model (of the *G. soja* factor), while a recessive model was deemed unlikely for two of the QTLs (sgA816 and sgK385). At

the remaining QTL (sgT153) neither dominance nor recessiveness could be deemed unlikely (Table 1).

Orthologous RFLP markers. To determine whether soybean shares orthologous QTLs with cowpea and mung bean, a set of 17 probes previously linked to seed weight QTLs in mung bean and cowpea (see Fatokun et al. 1992) were included in the genetic analysis of seed weight in the present study. Evidence for orthologous genes was shown by the colinearity of RFLP markers in genomic regions of soybean and cowpea that have significant effects on seed weight (Figure 4). Although this genomic region was not significant for seed weight in mung bean, colinearity of RFLP markers is maintained (Figure 4). Furthermore, a significant epistatic interaction between marker loci in this genomic region and unlinked seed weight QTLs is conserved in all three species (Table 2 and Fatokun et al. 1992). In soybean, a two-way ANOVA revealed that marker sgA816 significantly influenced a second unlinked locus (mgM185b). This suggests that when sgA816 is homozygous for the *G. soja* allele, it diminished the effects of the *G. max* allele on seed weight at mgM185b (Figure 5). Similarly, Fatokun et al. (1992) showed that the same marker locus sgA816 (or a nearby QTL) significantly influenced the expression of seed weight at a QTL near marker cgO103 in cowpea and marker mgM182 in mung bean in a similar manner. Interestingly, marker locus mgM185b is linked to cgO103

in cowpea and is located near a significant seed weight QTL in both cowpea and mung bean.

DISCUSSION

Seed weight analysis of the F_2 and $F_{2:3}$ generations showed a continuous distribution, indicative of a quantitative trait. The mean values of both generations were significantly closer to the value of the small seeded parent, suggesting partial dominance of alleles for small seed weight (Figure 1). This observation is consistent with previous reports in several plant species (Drabo et al. 1984; Vallejos and Chase 1991; Doebley et al. 1994). Nienhuis et al. (1987) reported a similarly skewed distribution that favored the alleles for the wild parent of an interspecific cross in tomato which they attributed to gametic selection. In the present study, the genotypic composition of the soybean population appears to be normally distributed with a mean *G. max* percentage of 0.475, which suggests that dominance and not gametic (or zygotic) selection is responsible for the smaller population means (Figure 6).

Employing both MapMaker-QTL and one-way ANOVA we identified three and five putative QTLs affecting seed weight in the F_2 and $F_{2:3}$ generations of this cross, respectively (Table 1). To reduce the probability of erroneously declaring a marker significant (Type I errors), we employed a strategy of mapping in two different

generations. Of the five QTLs mapped in the $F_{2:3}$ generation, two were significant ($P<0.01$) in the F_2 generation. The remaining three markers (sgA023, sgK384, sgT153) were significant at $P<0.05$ in the F_2 generation (Table 1). One marker locus (sgK385), highly significant in the F_2 generation, was not significant in the $F_{2:3}$ generation ($P=0.53$) and may have been erroneously declared significant in the F_2 generation (Type I error). Alternatively it may be that this putative QTL was significantly influenced by the environment (greenhouse verse field) in which it was tested. Changes in the importance of specific QTLs among different environments have been reported in tomato for fruit characteristics (Paterson et al. 1991). These results suggest that use of multiple generations provides a better basis with which to make conclusions about potentially useful QTLs.

The low level of variation (7.1 to 21.1%) explained by individual markers in these generations confirms the quantitative nature of seed weight inheritance (Table 1). For most of the QTLs identified, at least one of the three types of gene action tested (additive, dominant, or recessive), was rejected as unlikely. However, two modes of inheritance were never determined as unlikely. The inability to reject two modes of inheritance may be the result of the limited size of the population, but likely suggests that most of these QTLs exhibit partial dominance or recessiveness (Paterson et al. 1991). In addition to detecting major QTLs, we detected several highly significant epistatic interactions (Table 2). Previously, epistasis has been detected in mapping studies of soybean for hard seededness (Keim et al. 1990) and seed protein and oil

content (Lark et al. 1994) and in cowpea and mung bean for seed weight (Fatokun et al. 1992). The importance of these interactions is evidenced by their significant partial r^2 values (Table 2). The importance of epistasis in phenotype expression is supported by a large body of research in quantitative genetics (Allard 1988). In a combined analysis these markers and their significant interactions explained 50 and 60% of the variation for seed weight in the F_2 and $F_{2:3}$ generations, respectively. The magnitude of seed weight variation explained by these markers is substantial in view of the quantitative nature of the trait. The unexplained variation in this experiment may be attributed to the environment, colinearity in the data set with other traits influencing seed weight, measurement inaccuracies, and QTL or epistatic interactions with only small effects on seed weight.

Paterson et al. (1991), suggested that if distantly related species share orthologous genes, then one might contemplate "comparative QTL mapping" or the use of QTL mapping information from one species to predict the properties of other related species. Recently, Fatokun et al. (1992), provided RFLP evidence for the conservation of orthologous seed weight genes between two species of the genus *Vigna* (cowpea and mung bean). In the present study, we report evidence for conserved orthologous seed weight genes between two genera of the family Leguminosae. In soybean (*Glycine*; linkage group M) and cowpea (*Vigna*; linkage group vii), genomic regions significantly affecting seed weight are spanned by the same RFLP markers in the same linear order (Figure 4). Although a significant

QTL for seed weight was not detected for mung bean (*Vigna*; linkage group 2) in this genomic region, the linear order of the RFLP markers in this region are maintained (Figure 4). It may be that a seed weight QTL is present in this genomic region in mung bean but was not detected due to the small size of the F_2 populations studied by Fatokun et al. (1992). Furthermore, all three species had a highly significant epistatic interaction between a marker (sgA816) in this conserved region and RFLP marker loci at unlinked seed weight QTL(s).

Previously, Fatokun et al. (1992) reported the conservation of a single orthologous seed weight QTL on linkage group ii of cowpea and linkage group 1 of mung bean (according to current linkage map; Menancio-Hautea et al. 1993). In soybean (linkage group A), the colinearity of RFLP markers in this genomic region is maintained (data not presented), yet only slight evidence ($P=0.04$) for a seed weight QTL at this genomic region is detected. The lack of significant evidence for a seed weight QTL at this region in the soybean genome may indicate that i) no QTL for seed weight is present in this genomic region, ii) the importance of this region has been diluted during the evolution of soybean, or iii) this genomic region is not polymorphic in this cross (or perhaps in the species) and, therefore, is not detectable through standard RFLP mapping techniques.

These investigations have identified several seed weight QTLs in soybean. These QTLs have been described by genomic location, magnitude of effect, gene action (dominance and additive effects), and epistatic interactions. Furthermore, we have

identified a major orthologous seed weight QTL shared among three legume species. The implications of shared orthologous QTLs for important agronomic characteristics not only poses intriguing questions for quantitative genetics and evolutionary biology, but may provide answers for recalcitrant problems in plant breeding. If distantly related species share orthologous genes, it may be simpler to characterize complex genetic traits in well characterized genetic systems (e.g. Arabidopsis), and then apply that knowledge to related depauperate plant species. Currently we are investigating the genomic relationship of heterologous probes linked to fruit mass in tomato and kernel size in maize with seed weight QTLs in soybean.

Acknowledgements. This research was supported in part by grants from the Virginia Center for Innovative Technology, Montague Farms, Inc., and the Virginia Soybean Board.

REFERENCE

Akkaya MS, Bhagwat AA, Cregan PB (1992) Length polymorphisms of simple sequence repeat DNA in soybean. Genetics 132:1131-1139

Allard RW (1988) Genetic changes associated with the evolution of adaptedness in cultivated plants and their wild progenitors. J Hered 79:225-238

Burton JW (1987) Quantitative Genetics: Results relevant to soybean breeding. In Soybeans: Improvement, production and uses. American Society of Agronomy

Concibido V, Denny RL, Boutin SR, Hautea R, Orf JH, Young ND (1994) DNA marker analysis of loci underlying resistance to soybean cyst nematode (*Heterodera glycines* Ichinohe). Crop Sci 34:240-246

Drabo I, Redden R, Smithson JB, Aggarwal VD (1984) Inheritance of seed size in cowpea (*Vigna Unguiculata* (L.) Walp). Euphytica 33:929-934

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

Doebley J, Bacigalupo, Stec A (1994) Inheritance of kernel weight in two maize-teosinte hybrid populations for crop evolution. J Hered 85:191-195

Fatokun CA, Menancio-Hautea DI, Danesh D, Young ND (1992) Evidence for orthologous seed weight genes in cowpea and mung bean based on RFLP mapping. Genetics 132:841-846

Frey KJ, Horner T (1957) Heritability in standard units. Agron J 49:59-62

Griffis G, Wiedermann L (1990) Marketing food-quality soybeans in Japan. American Soybean Association, P.O. Box 27300, St. Louis, MO 63141

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

Kreike DM, de Koning JRA, Vinke JH, van Ooijen JW, Gebhardt C, Stiedema WJ (1993) Mapping of loci involved in quantitatively inherited resistance to the potato cyst nematode *Globodera rostochiensis* pathotype Ro1. Theor Appl Genet 87:464-470

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newberg L (1987) Mapmaker: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174-181

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

Lark KG, Orf J, Mansur LM (1994) Epistatic expression of quantitative trait loci (QTL) in soybean [*Glycine max* (L.) Merr.] determined by QTL association with RFLP alleles. Theor Appl Genet 88:486-489

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

Menancio-Hautea D, Fatokun CA, Kumar L, Danesh D, Young ND (1993) comparative genome analysis of mungbean (*Vigna radiata* L. Wilczek) and cowpea (*V. unguiculata* L. Walpers) using RFLP mapping data. Theor Appl Genet 86:797-810

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

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

Saghai Maroof MA, Soliman DM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci USA 81:8014-8018

Smartt J (1990) Grain legumes: Evolution and genetic resources. Cambridge university press, Cambridge

Smith JD, Kinman ML (1965) The use of parent-offspring regression as an estimator of heritability. Crop Sci 5:595-596

Stuber CW (1992) Biochemical and molecular markers in plant breeding. Plant Breed Rev 9:37-61.

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

Suiter KA, Wendel JF, Case JS (1983) Linkage-1: A Pascal computer program for the detection and analysis of genetic linkage. J Hered 74:203-204

Tanksley SD, Bernatzsky R, Lapitan NL, Prince JP (1988) Conservation of gene repetorire but not gene order in pepper and tomato. Proc Natl Acad Sci USA 85:6419-6423

Vallejos CE, Chase CD (1991) Linage between isozyme markers and a locus affecting seed size in *Phaseolus vulgaris* L. Theor Appl Genet 81:413-419

Williams JGK, Kubelik Ar, Livak KJ, Rafalsky A, Tingey SV (1990) DNA polymorphisms amplified by arbitary primers are useful as genetic markers. Nucleic Acids Res 18:6531-6535

Yu YG, Saghai Maroof MA, Buss GR, Maughan PJ, Tolin SA (1994) RFLP and microsatellite mapping of a gene for soybean mosaic virus resistance. Phytopathology 84:60-64

Zhang Q, Saghai Maroof MA, Kleinhofs A (1993) Comparative diversity analysis of RFLPS and isozymes within and among population of *Hordeum vulgare* ssp. *spontaneum*. Genetics 134:909-916

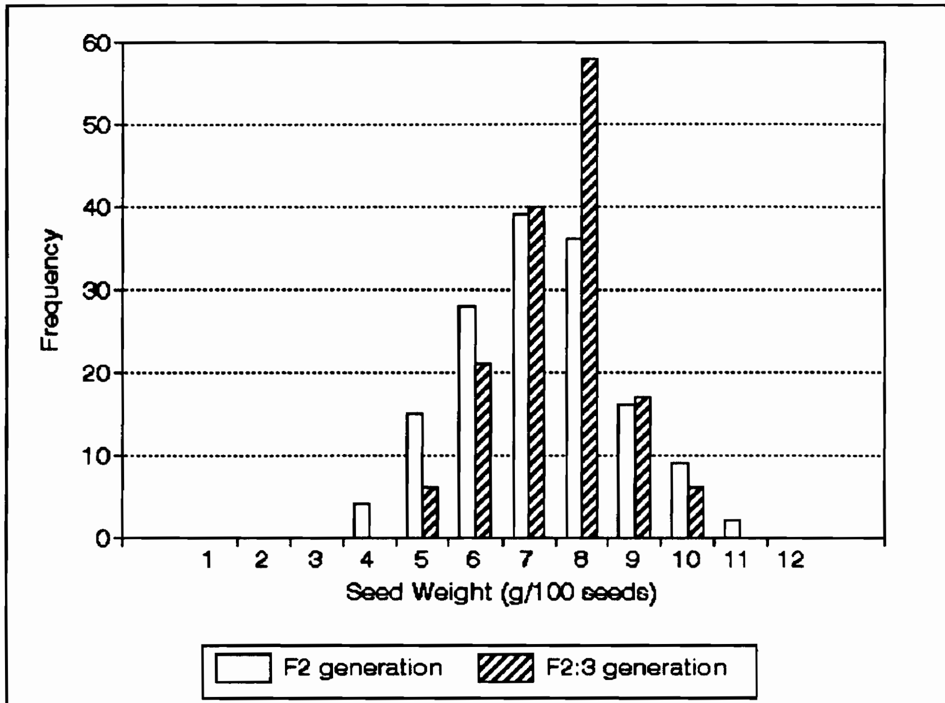


Figure 1: Frequency distribution for seed weight in the F_2 (open bars) and the $F_{2:3}$ generations (filled bars) of a cross between *G. max* and *G. soja*. Both distributions approximate a normal distribution. Mean seed weight and expected midparental values for the F_2 and $F_{2:3}$ generations were 6.7 and 7.1 g/100 seeds and 12.8 and 13.1 g/100, respectively.

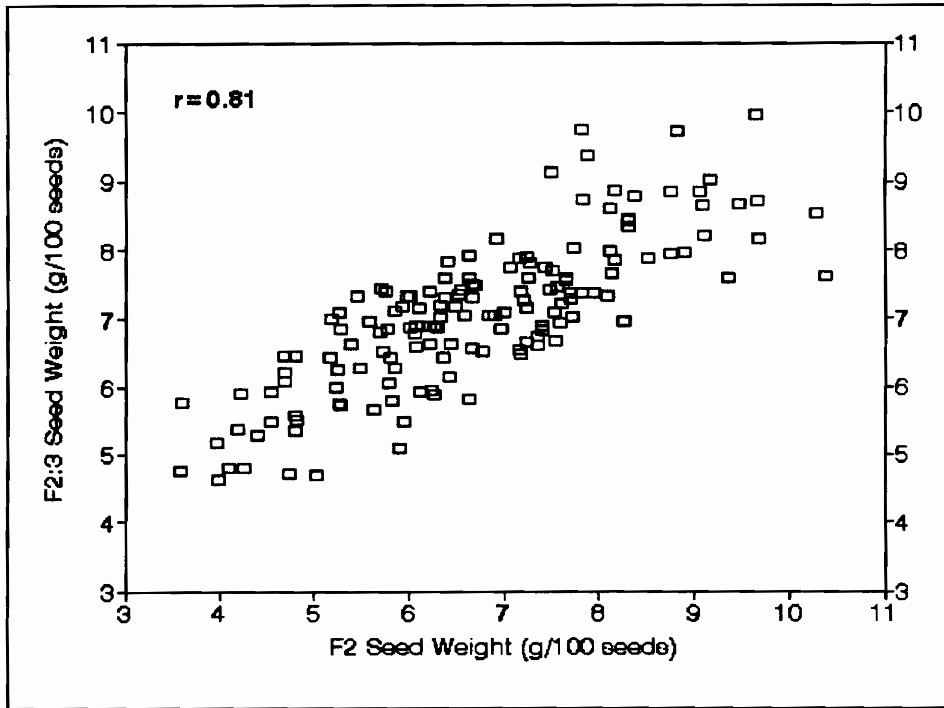


Figure 2: Regression plot of F_2 seed weight on $F_{2:3}$ seed weight. The correlation coefficient (r) is significant at the 0.001 level.

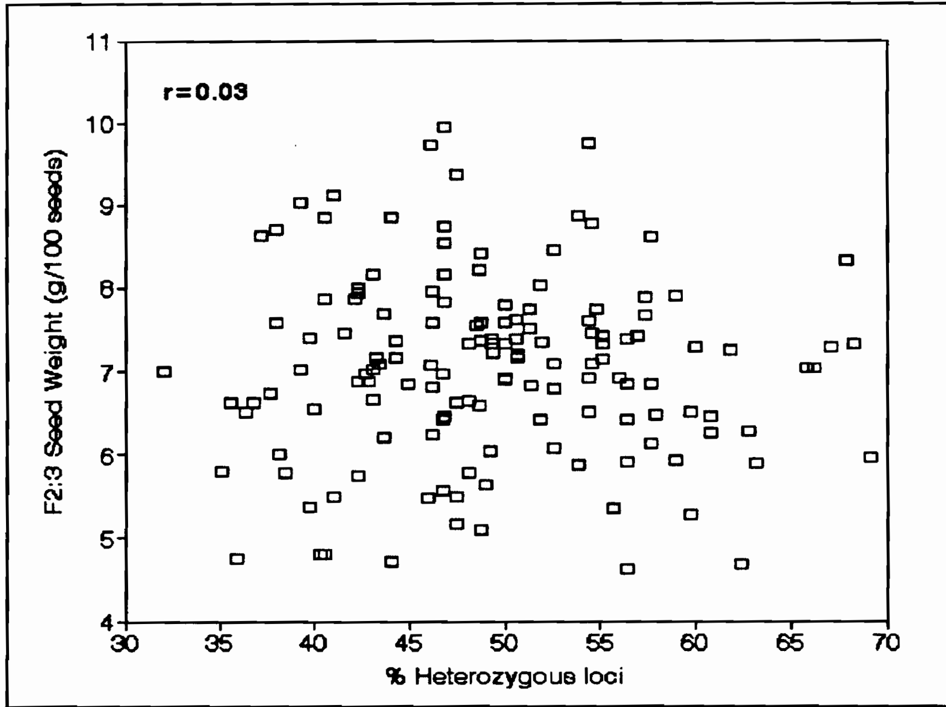


Figure 3: Regression plot showing the association of seed weight ($F_{2:3}$) with percent heterozygous loci. The correlation coefficient (r) is not significant.

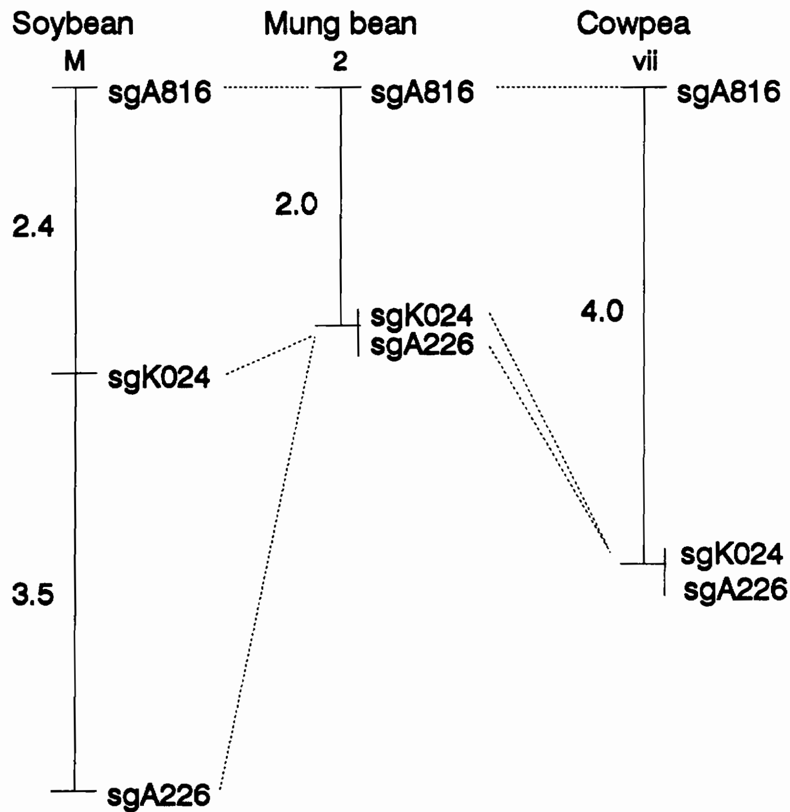


Figure 4: Orthologous genomic regions influencing seed weight in soybean, cowpea and mung bean. RFLP markers in this region are significantly associated with seed weight in soybean and cowpea. Epistatic interaction between marker (sgA816) and unlinked markers are significant in all three genomes (Table 2 and Figure 5). Colinearity of RFLP markers is maintained in all three genomes.

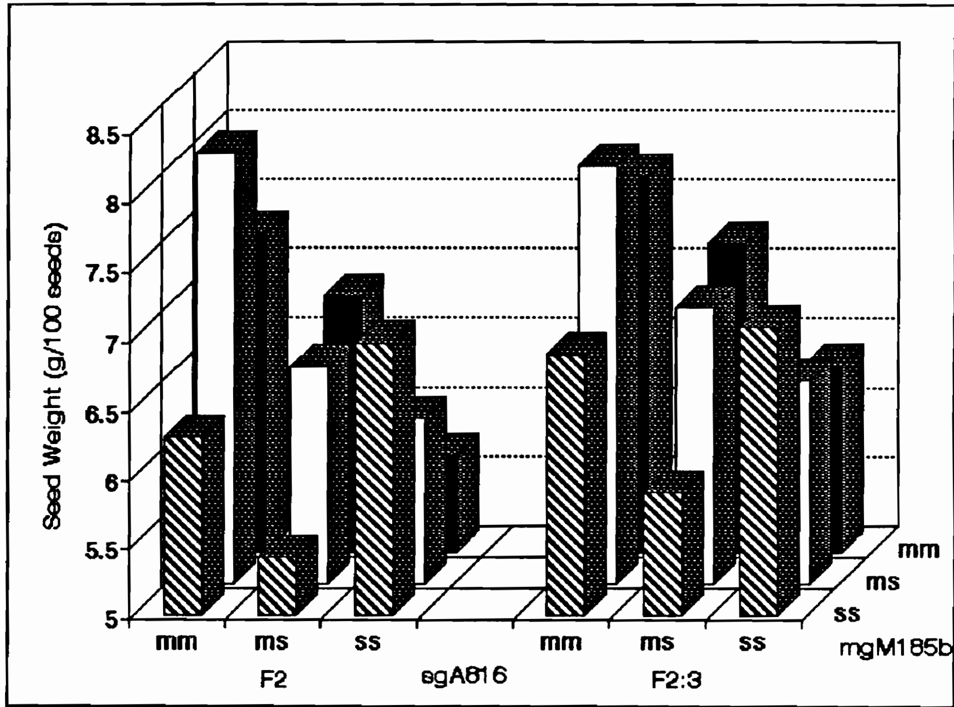


Figure 5: Epistatic interaction between marker loci sgA816 and mgM185b in the F_2 and $F_{2:3}$ generations of a cross between *G. max* and *G. soja*. Mean seed weights are plotted relative to their genotypic class at markers sgA816 and mgM185b. These results suggest that when pA816 is homozygous for the *G. soja* allele, it diminishes the effects of the *G. max* allele at mgM185b on seed weight.

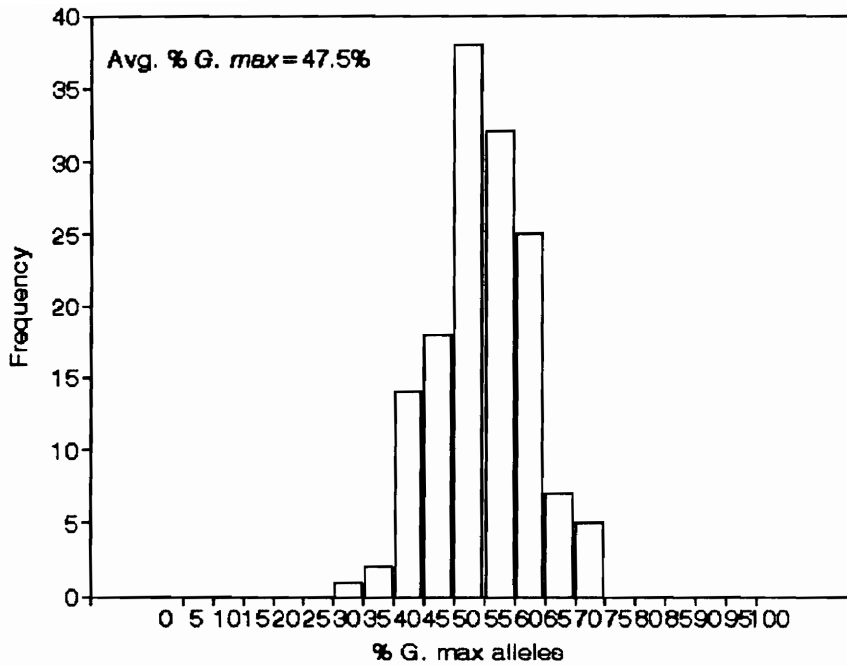


Figure 6: Genotypic composition of the F_2 individuals. Percent *G. max* was determined for each F_2 individual by dividing the total number of *G. max* alleles by the total number of alleles detected for all markers.

Table 1: Location, gene action and effect of markers significantly associated with seed size.

Marker	Gen[a]	LG[b]	MM[c]	MS	SS	%Var[d]	P>F[d]	Mode[e]
sgA118	F_2	B	7.7	6.6	5.8	21.1	0.001**	
	$F_{2:3}$		7.5	7.1	6.4	14.2	0.001**	AR
sgA816	F_2	M	7.3	6.5	6.4	7.8	0.004**	
	$F_{2:3}$		7.6	6.9	6.7	9.8	0.001**	DA
sgA023	F_2	L	7.1	6.8	6.3	4.3	0.049*	
	$F_{2:3}$		7.5	7.1	6.6	8.7	0.002**	AR
sgK384	F_2	J	7.1	6.8	6.2	5.7	0.017*	
	$F_{2:3}$		7.5	7.2	6.5	10.5	0.001**	AR
sgT153	F_2	A	7.1	6.8	6.2	4.9	0.032*	
	$F_{2:3}$		7.5	7.1	6.6	7.1	0.007**	ARD
sgK385	F_2	L	7.5	6.6	6.2	10.5	0.001**	AD
	$F_{2:3}$		7.2	7.0	6.9	0.9	0.535	

[a]The generation in which the QTL was detected (F_2 or $F_{2:3}$)

[b]Linkage group as designated in current USDA-ISU map

[c]MM-homozygous *G. max*; MS-heterozgous; SS-homozygous *G. soja*; measured as g/100 seeds

[d]Percent phenotypic variation explained by each marker locus and their corresponding P-values, from one-way ANOVA; **=significant at P>0.01; *=significant at P>0.05.

[e]The possible pure modes of gene action (Mode) for each QTL as detected in the $F_{2:3}$ generation (A-additive; D-dominant; R-recessive). The most likely mode is listed first.

Table 2: Significant interactions between genetic markers as determined using two-way ANOVA.

Interaction	Gen[a]	P>F	r^{2*}
sgA118*sgA426	F_2	0.008	4.8
	$F_{2:3}$	0.005	5.8
sgA816*sgM185b	F_2	0.002	7.5
	$F_{2:3}$	0.004	6.0
sgA118*sgA112	F_2	0.008	4.7
sgA118*sgA023	$F_{2:3}$	0.007	7.3

[a]The generation in which the interaction was detected (F_2 or $F_{2:3}$)

*Partial r^2

CHAPTER IV.

Molecular marker analysis of quantitative trait loci influencing seed quality characteristics in soybean (*Glycine max* L. Merr.)

ABSTRACT

The objectives of this study were to use molecular markers (RFLPs, SSRs, and RAPDs) and interval mapping techniques to position and characterize quantitative trait loci (QTL) controlling seed quality in soybeans. The seed quality traits analyzed in this study included seed weight, protein, oil, sucrose, and calcium. An F_2 population was constructed from an interspecific cross between a *Glycine max* breeding line (V71-370) and a *G. soja* plant introduction (PI407.162). Quantitative traits were analyzed in 150 $F_{2:3}$ lines grown in a replicated field experiment. QTLs significantly associated with these seed quality components were determined using the computer program MapMaker-QTL. Two QTLs were detected for protein and calcium content, five for oil content and seed weight and six for sucrose content. Percent phenotypic variation explained at individual QTLs varied from 6.6 to 34%. The total phenotypic variation explained by all QTLs for specific traits were 42.5, 36.7, 49.0, 53.1, and 42.6% for seed weight, protein, oil, sucrose, and calcium, respectively. One RAPD marker accounted for 34% of the total phenotypic variation associated with seed protein content. Seed quality traits tended to share genomic regions (or QTLs) of importance. Of the 11 genomic intervals identified, six were associated with more than one trait. Genetic correlations between seed quality traits and shared genomic intervals are discussed.

Key words: Genetic mapping - Quantitative trait loci - Soybean - Seed quality - Restriction fragment length polymorphism

INTRODUCTION

Soybean seed can be processed for protein and oil or utilized directly as a raw product in the production of several oriental soyfood items. While only a small percentage of the total soybean seed produced is used directly for manufacturing soyfoods, the demand for "food quality" soybeans is increasing at a rate of 3 to 5 percent per annum (Griffis and Wiedermann 1990). Because most U.S. soybean varieties are not suitable for soyfood production (due to special requirements of appearance, taste, and nutrition), producers are offered premiums to produce specialty varieties for this market (Carter 1987). Tofu, natto, and miso (traditional Japanese soyfoods) comprise nearly 90% of the food quality soybean market (Taira 1990). The quality and yield of these soyfoods are greatly effected by the characteristics of the soybean seed utilized to produce them. Large seeded soybeans (>20 mg/seed) with high protein and sucrose content and low oil and calcium content are preferred for tofu and miso processing, while small seeded (<8 mg/seed) soybean varieties with high sucrose and low calcium content are preferred for natto production (Griffis and Wiederman 1990; Taira 1990).

Increasing interest in developing soybean varieties that fit into this specific market has made breeding for food quality soybeans a major objective of some breeding programs. Breeding concerns for these specialty soybeans encompass several quantitatively inherited seed quality traits including, protein, oil, carbohydrate, and mineral content as well as seed weight and appearance. While soybean breeders

have been successful in manipulating many of these traits, their inheritance and underlying genetic control have not been fully elucidated. Knowledge of the genetic basis for these seed characteristics should help breeders accelerate the development of specialty varieties destined for food quality purposes.

Recently, the study of quantitatively inherited traits has been greatly facilitated by the advent of new genetic markers and mapping technologies, most notably restriction fragment length polymorphisms (or RFLP markers) and interval mapping techniques. Interval mapping of quantitative trait loci using genetic linkage maps provides an analytical method that positions putative quantitative trait loci (QTLs) between pairs of qualitative genetic markers, e.g., RFLPs (Lander and Botstein 1989). Utilizing this approach has several advantages over the traditional method of associating QTL effects with single genetic markers using one-way analysis of variance, including: i) highly efficient detection of QTLs with a limited false-positive error rate, ii) unbiased estimates of phenotypic effects at QTLs, and iii) positioning of QTLs between discrete genetic markers. Furthermore, the development of a computer program, Mapmaker-QTL, (Lincoln and Lander 1990) greatly simplifies the use of interval mapping in a variety of quantitative genetic studies.

The systematic dissection of quantitative traits using molecular markers and interval mapping has been demonstrated for a variety of plant species, including fruit characteristics in tomato (Patterson et al. 1991), agronomic traits in soybean (Mansur et al. 1993), heterosis in maize (Stuber 1992). and quantitative disease resistance in

potato (Kreike et al. 1993). Information on the genomic location, gene number, gene action and the association of genetic markers with putative QTL(s) for seed quality characteristics should help breeders to develop reliable selection methods and accelerate the construction of beneficial allelic combinations for multiple seed quality traits in elite soybean varieties.

The objectives of this research were to use molecular markers and interval mapping techniques to position and characterize quantitative trait loci controlling protein, oil, sucrose, and calcium content as well as seed weight in soybean.

MATERIALS AND METHODS

Genetic Materials. The experimental population utilized in this study has been previously described by Maughan et al. (1994). Briefly, an F_2 population consisting of 150 individual plants was developed from an interspecific cross between an adapted large-seeded *G. max* breeding line (V71-370) and a small-seeded *G. soja* plant introduction (PI407.162). At maturity, each plant was harvested and threshed separately to produce 150 F_2 derived F_3 ($F_{2:3}$) lines. All 150 $F_{2:3}$ lines were then planted in a randomized complete block design with two replications at the Virginia Crop Improvement Association Foundation Seed farm near Mt. Holly, Va. When individual $F_{2:3}$ lines reached maturity, each was harvested and threshed separately to produce $F_{2:4}$ lines (F_4 seed). F_2 genotypes for three morphological loci (pubescence

color, T/t; pod color, L1/l1; and pubescence type, Pa1/pa1) were determined by 7screening $F_{2:3}$ progeny.

Quantitative trait analysis. Five seed quality traits were analyzed using bulk samples of F_4 seed from each $F_{2:3}$ line. These traits included protein, oil, sucrose, and calcium content as well as seed weight. Seed protein and oil content were determined from 10 g samples of ground seed from each $F_{2:3}$ line using Kjeldahl procedures according to American Association of Cereal Chemist (AACC) guidelines. Protein and oil analyses were performed at the USDA Northern Regional Research Center, Peroia, Il. Seed sucrose content was determined using a sucrose/D-glucose enzymatic assay kit (Boehringer Manneheim, Indianapolis, IN; Cat. No. 139 041) with the following modifications. Briefly, a 1 g ground soybean sample for each $F_{2:3}$ line was prepared for the enzymatic sucrose assay by extracting total sugars with 100 ml of 80% ethanol for 20 hrs at 23 °C. After extraction, a 1 ml sample was forced through a 0.45 μm filter (Acrodisc LC PVDF; Gelman) and evaporated to dryness using a Reactitherm heating/evaporation unit set at 98 °C. Sugar extracts were then redissolved in 1 ml distilled deionized water and assayed for sucrose content using an invertase/hexokinase enzymatic assay according to the manufacturer guidelines. Calcium content was determined from 5 g ground samples using an atomic absorption spectrophotometer. Calcium content assays were made at the Tidewater Agricultural Research and Extension Center, Suffolk, Va. Seed weight was

determined by weighing a random sample of 100 seeds from each $F_{2:3}$ plot. Seed weight is reported in g/100 seeds. Protein, oil, sucrose, and calcium values are reported on a dry weight basis as percent of dry matter.

Molecular marker assay. DNA extraction, Southern blotting, and hybridization procedures used were as described elsewhere (Saghai Maroof et al. 1984; Zhang et al. 1993). One hundred and eighty-seven low copy clones, derived from three *Pst*I genomic libraries (obtained from soybean, mung bean, and cowpea), were screened for polymorphism in the parents using five restriction enzymes (*BamH*I, *Dra*I, *EcoR*I, *EcoR*V, and *Hind*III). These clones were selected based on their position in the current USDA-ISU soybean genetic map to provide maximum genome coverage (Diers et al. 1992) or based on previous reports of linkage to putative seed weight QTLs in cowpea and/or mung bean (Fatokun et al. 1992). DNA clones were generously provided by R.C. Shoemaker (Iowa State University, Ames, IA) and N.D. Young (University of Minnesota, St. Paul, MN). A total of 84 RFLP clones were polymorphic between the parents of this cross. Clones derived from soybean, cowpea, and mung bean libraries will be designated with the prefix sg, cg, or mg, respectively.

In addition to screening for RFLPs, a set of forty randomly amplified polymorphic DNA (RAPD) primer (Operon primer kits B and O; Alameda, CA) and eight simple sequence repeats (SSRs or microsatellites) were screened for

polymorphism according to the procedures outlined by Williams et al (1990) and Yu et al. (1994), respectively. Of the 40 RAPD Primers screened, three produced reproducible polymorphic markers. RAPD markers will be given the prefix R, followed by the Operon Kit designation (B or O), followed by the primer number. Of the eight SSR markers screened, four have been described by Akkaya et al. (1992) and four were described by Yu et al. (1994). Four SSR markers were polymorphic between the parents of this cross. SSR markers will be given the prefix sm. In total, genotypes for 92 genetic loci were scored for 150 F_2 individuals.

Interval mapping and QTL analysis. Trait means, standard deviations, correlations, and normality were determined using the computer program SAS (Statistical Analysis Systems, Cary, NC). Most probable map order and distances were determined by multi-point linkage analysis using the computer program MapMaker 2.0 at LOD=2.4 and with a maximum Haldane distance of 50 centiMorgans (cM) (Lander et al. 1987). Interval mapping of molecular marker-QTL associations was performed using the computer program MapMaker-QTL 1.1b (Landers and Botstein 1989) for 77 linked markers. Non-linked markers were analyzed by genotypic class means using one-way analysis of variance. Cosegregation of genetic markers and putative QTLs was inferred whenever LOD (logarithm to the base 10 of the likelihood odds ratio) scores in MapMaker-QTL exceeded 2.0 or when ANOVA F-test indicated a significance level of $P<0.01$. Interval mapping results obtained from MapMaker-QTL were

confirmed by one-way analysis of variance. Gene action at significant QTLs was evaluated by comparing the fit of three Mendelian models using MapMaker-QTL (i.e., either additive, dominant, or recessive; Lincoln and Lander 1990). A LOD reduction of one in likelihood of the presence of a QTL in a specific interval was inferred as evidence that a given type of gene action was unlikely. Only the modes of inheritance that could not be rejected are reported in Table 3.

RESULTS

Phenotypic analysis. Seed quality trait means, ranges, and standard deviations measured for $F_{2:3}$ lines are presented in Table 1. Significant genetic variation was observed among the $F_{2:3}$ lines for all characteristics studied. Transgressive segregation was observed for three (protein, oil, and calcium content) of the five traits. Significant deviation from expected midparental values was observed only for seed weight (13.6 g/100 seed). Figure 1 presents the frequency distributions for seed weight, protein, oil, sucrose, calcium, and ln(calcium). All of the traits studied exhibited continuous variation, suggesting multigenic inheritance. Protein, oil, sucrose, and seed weight approximated normal distributions ($P<0.05$), while calcium content was skewed toward the lower value of the *G. max* parent (V71-370). The transformation, ln(Calcium%), improved normality, and was used for all analyses (Figure 1).

Correlations among seed weight, protein, oil, sucrose, and calcium content are presented in Table 2. Consistent with previous reports (Burton and Brim 1981; Hymowitz et al. 1972), high seed protein was correlated with low seed oil (r=-0.63) and low seed sucrose (r=-0.46), while high seed oil was correlated with high sucrose content (r=0.45). Seed weight was positively correlated with seed oil (r=0.42) and sucrose (r=0.52) and negatively correlated with calcium content (r=-0.31). A slight negative correlation observed between seed weight and seed protein was not significant (r=-0.12). Calcium content was negatively correlated with seed oil (r=-0.31) and sucrose (r=-0.38) content. No correlation was observed between protein and calcium content (r=0.05). While correlation values between several seed quality characteristics are highly significant, it should be noted that considerable variation (r^2) for all traits remains unexplained.

Interval mapping of seed quality QTLs. QTLs controlling seed protein, oil, sucrose, calcium content, and seed weight were localized to 11 genomic intervals in the soybean linkage map (Table 3). The number of putative QTLs detected varied greatly for each trait. Two QTLs were detected for protein and calcium content, five QTLs were detected for oil content and seed weight and six QTLs were detected for sucrose content. Percent phenotypic variation explained at individual QTLs varied dramatically from a low of 6.6% for seed oil content (sgA23-L1) to a high of 34.0% (RB4b-sgK227) for protein content (Table 3). By simultaneously fitting all significant

QTLs for a specific trait to a single genetic model, we were able to determine the total amount of phenotypic variation explained by the QTLs identified. The percent of the total phenotypic variation explained by all QTLs for a specific trait were 42.5, 36.7, 49.0, 53.1, and 42.6% for seed weight, protein, oil, sucrose, and calcium, respectively. The unexplained variation in these analyses may be attributed to the environment, colinearity in the data set, measurement inaccuracies, epistatic interaction, or QTL(s) with undectable effects. None of the 14 unlinked loci, tested with one-way ANOVA, were significantly associated with variation for any of the traits studies.

The likely mode of inheritance (gene action) of each QTL was investigated by comparing the fit of individual QTLs to three Mendelian models using MapMaker-QTL (i.e., either dominant, recessive, or additive). In these analyses, the alleles of the *G. max* parent (V71-370) were arbitrarily considered as dominant alleles while the alleles of the *G. soja* parent (PI407.162) were considered recessive. The additive effects, dominance effects, and likely modes of inheritance for each QTL are presented in Table 3. For most of the QTLs identified in this study, at least one type of gene action was rejected. In only a few cases, were two modes of inheritance determined to be unlikely, suggesting that partial dominance (or recessiveness) is the major mode of gene action exhibited by these QTLs. The direction of the effects of all QTLs for seed weight, oil, and sucrose were such that the *G. max* alleles acted to increase seed weight, sucrose, and oil. This is not surprising given that the *G. max*

parent (V71-370) has significantly larger seed weight and higher levels of sucrose and oil content. However, for protein and calcium content, many QTLs displayed effects with unexpected directions, e.g., QTLs from the low protein parent (V71-370) contributed positively to protein content. Similar results for QTLs having unexpected effects have been reported for several maize traits (Doebley and Stec 1991).

The most striking feature of these analyses was the tendency of seed quality traits to share genomic regions (or QTLs) of importance. For example, the interval RB4b-sgK227 (spanning 7.7 cM) is significantly associated with phenotypic variation for protein (LOD=9.7), oil (LOD=4.7), and sucrose (LOD=4.1) content (Table 3 and Figure 2). Of the 11 genomic intervals identified, five were associated with a single trait, four were associated with two traits, and two were associated with three traits (Table 3). This clustering of important QTLs for related traits was also observed for developmental and morphological traits in soybean (Keim et al. 1990; Mansur et al. 1993).

DISCUSSION

In the genetic analysis of this interspecific cross, several potential seed quality QTLs were identified. The low level of variation (4.6 to 34%; Table 3) explained by the majority of individual QTLs and the continuous distribution formed by each trait (Figure 1) confirms the quantitative nature of inheritance for these seed quality

characteristics. Interestingly, in two of the five traits analyzed (protein and calcium), single QTLs explaining greater than 20% of the total phenotypic variation were detected. The level of variation explained by these QTLs is substantial in view of the quantitative nature of these traits. Such QTLs could provide an important tool for the rapid improvement of these characters through marker-assisted breeding efforts. However, because several researchers have reported significant changes in the importance of specific QTLs among populations and environments (Schön et al. 1994; Tanksley and Hewitt 1988), "major" QTLs need to be studied in several genetic backgrounds before marker-assisted selection strategies are implemented.

The tendency for several QTLs affecting different seed quality traits to share common genomic regions (intervals) occurred more frequently than would be expected by chance alone (Table 3 and Figure 2). This observation suggests that correlations between seed quality traits (Table 2), especially strong correlations, may be the result of pleiotropy at some important QTLs. This phenomenon may be explained by a key metabolic step which interrelates seed quality characteristics. For example, a major QTL controlling carbon partitioning during seed development could control the various levels of all major seed constituents. Alternatively it may be that different QTLs conditioning these traits are inherited in clusters as tightly linked loci. Similarly clustered QTLs for correlated traits have been observed for developmental and morphological traits in soybean (Mansur et al. 1993; Keim et al. 1990) and for fruit characteristics in tomato (Patterson et al. 1991). The potentially pleotrophic

nature and the observed correlations between seed quality traits has direct implications for breeding of speciality soybean cultivars. For example, the positive correlation (r=0.52) between seed size and sucrose content will make breeding for small seeded soybean cultivars with high sucrose content more difficult. Indeed, in the population examined in this study, no $F_{2:3}$ lines with small seed size and high sucrose content were observed. Although the majority of seed quality traits measured in this study showed highly significant correlations with at least one other seed quality trait, the level of variation explained by specific correlations (r^2) are generally low, indicating that for most combinations of traits, breeding progress is possible.

To investigate the similarity of QTLs for protein and oil content in different genetic backgrounds, we compared our results with those reported by Diers et al. (1992) and Mansur et al. (1993). Comparison of QTL mapping results between our study and that of Diers et al. (1992), for a different interspecific soybean population, identified a major QTL for both protein and oil in the same region (RB4b-sgK227) of linkage group I (USDA-ISU soybean genetic map, January, 1992). This genomic region explained 42 and 28% of the variation for protein and oil content in the study by Diers et al. (1992) and explained 34 and 19% of the variation for the protein and oil in our analysis. Furthermore, a second genomic region (sgA86-sgA963) significantly associated with oil content was maintained on linkage group E in both studies. Conserved genomic locations for several QTL across several genetic

backgrounds has been reported for yield in maize (Helentjaris 1993), fruit characteristics in tomato (Paterson et al. 1991) and for agronomic traits in soybean (Mansur et al. 1993). Interestingly, however, neither of these regions was detected in the intraspecific (*G. max* X *G. max*) population analyzed by Mansur et al. (1993). Indeed, several QTLs detected by our analysis were not detected by Mansur et al. (1993) or Diers et al. (1992) and vice versa. These results suggest that the importance of specific QTLs for protein and oil content (and perhaps other quantitative traits) may change from population to population and/or the genomic location of major QTLs may not be conserved between interspecific and intraspecific crosses in soybean. Such observations re-enforce the need to investigate complex genetic characteristics in several different genetic backgrounds (Paterson et al. 1991; Mansur et al. 1993).

Interval mapping with molecular markers provides a powerful means for the identification of quantitative trait loci in soybean. The simultaneous evaluation of related characteristics provides useful insight on the genetic basis of correlated traits and critical information necessary for breeders striving to manipulate these components. In the present study, we identified several highly significant genomic intervals associated with several seed quality traits in soybean, including seed weight, protein, oil, sucrose, and calcium content. These QTLs have been described by genomic location, magnitude of effect, and gene action (additive, dominant, recessive). Furthermore, we have shown a genetic basis for the observed correlations

among many of these seed quality traits. Indeed, greater than 50% of the genomic intervals identified in this study, significantly influenced the expression of two or more seed quality traits. Whether the action of these genomic regions is the result of a pleiotropic effect of a single QTL or the result of a cluster of tightly linked QTLs for several different seed quality traits is unclear. Further research in this area should provide useful insight not only for plant breeding and plant physiology, but for developmental genetics as a whole.

Acknowledgements. The authors acknowledge the work of Joe Parrish for his assistance in collecting sucrose data. This work was supported, in part, by grants from the Virginia Center for Innovative Technology, Montague Farms, Inc, and the Virginia Soybean Board.

REFERENCES

Akkaya MS, Bhagwat AA, Cregan PB (1992) Length polymorphisms of simple sequence repeat DNA in soybean. Genetics 132:1131-1139

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

Carter T (1987) Genetic alteration of seed size: breeding strategies and market potential. p. 33-45. In: Wilkinson, D. (ed.). Proceedings of the Seventeenth Soybean Seed Research Conference. American Seed Trade Assoc Washington, DC

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

Doebley J, Stec A (1991) Genetic analysis of the morphological differences between maize and teosinte. Genetics 129:285-295

Fatokun CA, Menancio-Hautea DI, Danesh D, Young ND (1992) Evidence for orthologous seed weight genes in cowpea and mung bean based on RFLP mapping. Genetics 132:841-846

Griffis G, Wiedermann L (1990) Marketing food-quality soybeans in Japan. American Soybean Association, P.O. Box 27300, St. Louis, MO 63141

Helentjaris TG (1993) RFLP analyses for manipulating agronomic traits in plants. In: Stalker HT, Murphy JP (eds) Plant Breeding in the 1990s. Redwood Ress Ltd., Melksham, UK

Hymowitz T, Collins FI Panczner J, Walker WM (1972) Relationship between the content of oil, protein, and sugar in soybean seed. Agron J 66:613-616

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

Kreike DM, de Koning JRA, Vinke JH, van Ooijen JW, Gebhardt C, Stiedema WJ (1993) Mapping of loci involved in quantitatively inherited resistance to the potato cyst nematode *Globodera rostochiensis* pathotype Ro1. Theor Appl Genet 87:464-470

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

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newberg L (1987) Mapmaker: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174-181

Lincoln SE, Lander ES (1990) Mapping genes controlling quantitative traits using MAPMAKER/QTL. A White head Institute for Biomedical Research Technical Report. January, 1990

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

Maughan PJ, Saghai Maroof MA, Buss GR (1994) Molecular marker analysis of seed weight: Chromosomal locations, gene action, and evidence for orthologous evolution among three legume species. PhD dissertation, Virginia Polytechnic Institute and State University, Blacksburg, Virginia

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

Saghai Maroof MA, Soliman DM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location and population dynamics. Proc Natl Acad Sci USA 81:8014-8018

Schön CC, Melchinger AE, Boppenmaier J, Brunklaus-Jung E, Herrmann RG, Seitzer JF (1994) RFLP mapping in maize: quantitative trait loci affecting testcross performance of elite european flint lines. Crop Sci 34:378-389

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

Taira H (1990) Quality of soybean for processed foods in Japan. JARQ 24:224-230

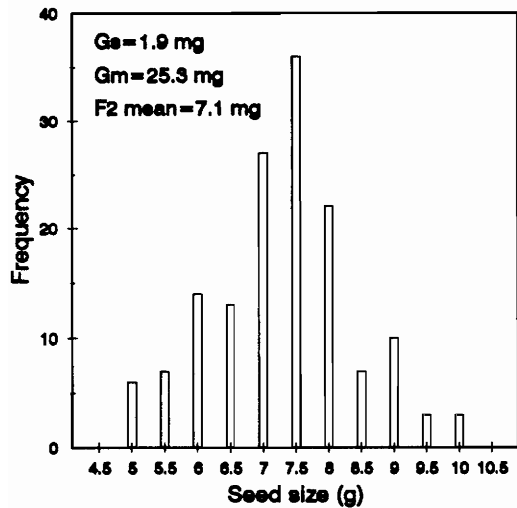
Tanksley SD, Hewitt J (1988) Use of molecular markers in breeding for soluble solids content in tomato-a re-examination. Theor Appl Genet 75:811-823

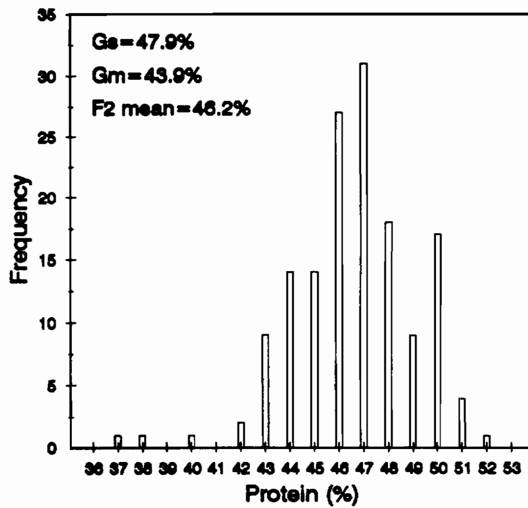
Williams JGK, Kubelik AR, Livak KJ, Rafalsky A, Tingey SV (19990) DNA polymorphism amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18:6531-6535

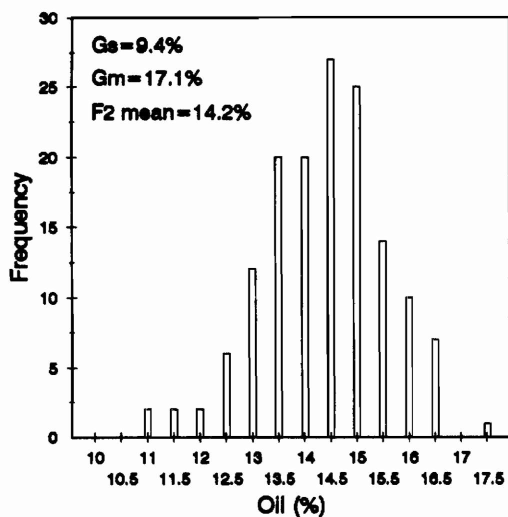
Yu YG, Saghai Maroof MA, Buss GR, Maughan PJ, Tolin SA (1994) RFLP and microsatellite mapping of a gene for soybean mosaic virus resistance. Phytopathology 84:60-64

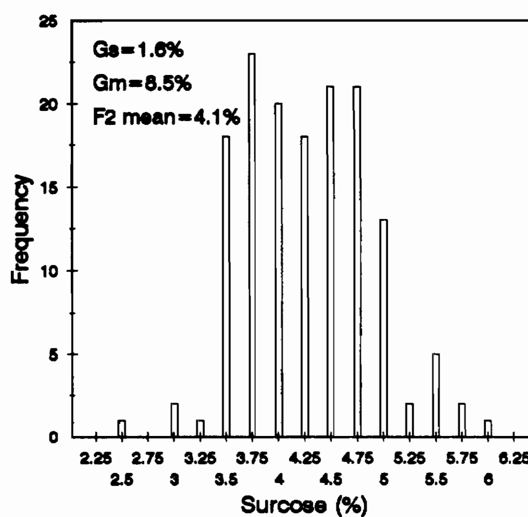
Zhang Q, Saghai Maroof MA, Kleinhofs A (1993) Comparative diversity analysis of RFLPs and isozymes within and among populations of *Hordeum vulgare* ssp. *spontaneum*. Genetics 134:909-916

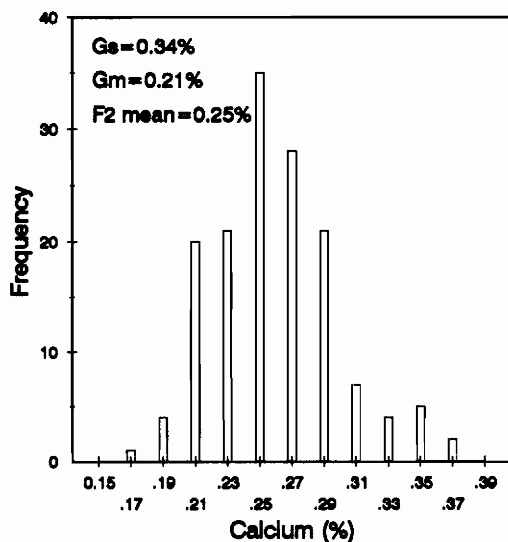
Figure 1: Frequency distributions for seed weight, protein, oil, sucrose, calcium, and ln(calcium). The F_2 distributions are approximately normal for seed weight, protein, oil, and surcose, but were skewed for seed calcium. Ln(calcium%) was used to improve normality. Parental and F_2 means for each trait are included (Gs=PI407.162 and Gm=V71-370).

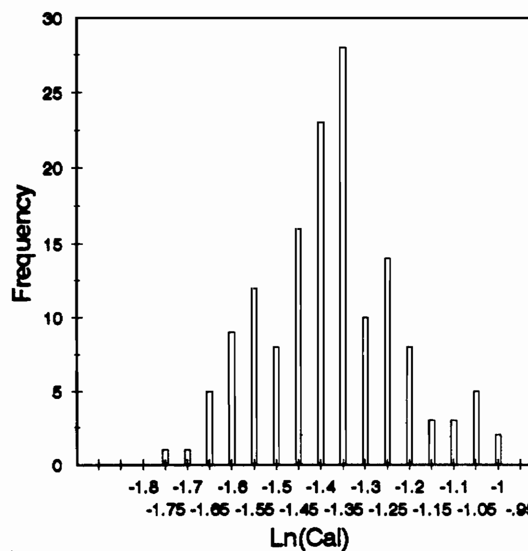

Gs=1.9 mg
Gm=25.3 mg
F2 mean=7.1 mg
Frequency
Seed size (g)


Gs=47.9%
Gm=43.9%
F2 mean=46.2%
Frequency
Protein (%)


Gs=9.4%
Gm=17.1%
F2 mean=14.2%
Frequency
Oil (%)


Gs=1.6%
Gm=8.5%
F2 mean=4.1%
Frequency
Surcose (%)


Gs=0.34%
Gm=0.21%
F2 mean=0.25%
Frequency
Calcium (%)


Frequency
Ln(Cal)

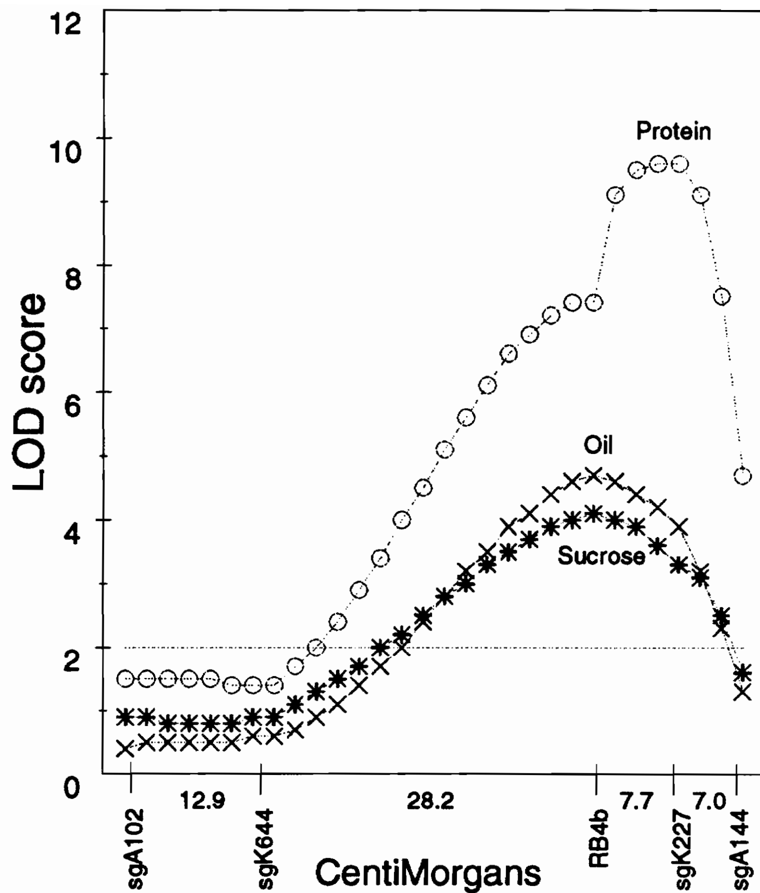


Figure 2: QTL likelihood plot for a genomic region influencing seed protein, oil, and sucrose. The interval RB4b-sgK227 is significantly associated with phenotypic variation for seed protein (LOD=9.7), oil (LOD=4.7), and sucrose (LOD=4.1). Distances between marker loci were determined using the computer program MapMaker and are reported in CentiMorgans. A LOD score of two was used as a significance threshold.

Table 1: Mean, standard deviation, range, and parental values for five seed quality traits measured in $F_{2:3}$ families.

Trait	N	Mean	SD[a]	Min	Max	*Gm*[b]	*Gs*
Seed size (mg)	148	7.1	1.1	4.6	10.0	1.9	25.3
Protein (%)	148	46.2	2.3	37.6	51.6	43.9	47.9
Oil (%)	149	14.2	1.3	10.7	20.4	17.1	9.4
Sucrose (%)	148	4.1	0.6	2.4	5.8	8.5	1.6
Calcium (%)	148	0.25	0.15	0.17	0.36	0.21	0.34

[a]Standard deviations

[b]*Gm*=V71.370; *Gs*=PI407.162

Table 2: Pearson correlation coefficients among five seed quality traits measured in $F_{2:3}$ families.

Trait	Protein	Oil	Sucrose	Calcium
Seed size	-0.12^{NS}	0.42^{**}	0.52^{**}	-0.31^{**}
Protein		-0.63^{**}	-0.46^{**}	0.05^{NS}
Oil			0.45^{**}	-0.31^{**}
Sucrose				-0.38^{**}

NS,*,denote not significant, significant at P<0.05 and significant at P<0.01, respectively.**

Table 3: Location, gene action, and effect of significant intervals containing quantitative trait loci (QTL) for seed quality characteristics.

Marker Interval[a]	LG[b]	Length (cM)	Additive Effect[c]	Dominance Effect[c]	Mode[d]	%Var	LOD
Seed Size							
sgA816-sgA226	M	2.4	-0.42	-0.20	DA	9.2	3.0
sgA118-sgA36	B	0.4	-0.51	0.18	AR	12.4	4.1
sgA23-L1[e]	L	17.2	-0.43	0.07	AR	8.6	2.8
sgK384-sgB122	J	17.4	-0.55	0.11	AR	12.6	3.5
sgT153-sgA136	A	29.1	-0.48	0.02	ARD	9.5	2.3
Protein							
sgK24-sgK226	M	36.0	-0.65	1.02	R	9.1	2.4
RB4b-sgK227	I	7.7	2.06	0.01	A	34.0	9.7
Oil							
T[e]-sgK397	C	7.0	0.53	0.58	D	12.4	4.3
sgA23-L1	L	17.2	-0.42	-0.29	DA	6.6	2.1
sgK384-sgB122	J	17.4	-0.51	0.33	RA	9.7	2.7
sgA86-sgA963	E	21.4	-0.41	0.41	RA	8.3	2.4
RB4b-sgK227	I	7.7	-0.81	-0.18	AD	18.6	4.7
Sucrose							
sgA186-sgA806a	F	18.9	-0.25	0.08	AR	6.9	2.3
sgA487-sgT169	A	4.7	-0.20	-0.11	DA	7.2	2.3
sgA23-L1	L	17.2	-0.25	-0.08	AD	9.4	3.1
sgA86-sgA963	E	21.4	-0.25	-0.12	DA	9.6	2.4
RB4b-sgK227	I	7.7	-0.36	0.01	A	15.1	4.1
sgT153-sgA136	A	29.1	-0.26	0.18	RA	12.1	2.7
Calcium[f]							
T-sgK397	C	7.0	-0.09	-0.05	DA	18.1	6.4
sgT153-sgA136	A	29.1	0.09	-0.05	RA	20.8	7.2

[a]Intervals were computed using the free genetic model by MapMaker-QTL with a significant LOD threshold of 2.0 or higher

[b]Linkage groups are as designated in current USDA-ISU map

[c]Additive and dominance effects are given for the QTL positions corresponding to the maximum LOD score for each significant interval

[d]The possible modes of gene action for each QTL (A-additive; D-dominant; R-recessive); the most likely mode is listed first

[e]L1-pubescence type; T-pubescence color

[f]%Calcium data was transformed to ln(cal) in order to satisfy normality assumptions

VITA

Peter Jeffrey Maughan was born Feburary 20, 1966 in Logan, Utah. Mr. Maughan attended Seoul American High School in Seoul, Korea, where he graduated in July, 1984. He then attended Brigham Young University, Provo, Utah, where he studied crop science. In May, 1989, he completed his studies and recieved a Bachelor of Science degree in Agronomy. He decided to continue his education at Brigham Young University, studing molecular genetics in the Department of Agronomy and Horticulture. In August, 1991, he received a Masters of Science degree. Also in 1991, he accepted an assistantship at Virginia Polytechnic Institute and State University to participate in a molecular genetics research project in the Department of Crop and Soil Environmental Sciences. He was admitted to the interdepartmental program of Genetics in 1993. Mr. Maughan is an elected member of the honor society of Gamma Sigma Delta. His other affiliations include the Genetic Society of America, American Society of Agronomy, Crop Science Society of America, and the American Association for the Advancement of Science.