SOYBEAN IMPROVEMENT OF SEED TRAITS AND RESISTANCE TO BUD BLIGHT VIA MOLECULAR MARKERS AND CLASSICAL METHODS

by

VASILIA APOSTOLOS FASOULA

(Under the direction of H. Roger Boerma)

ABSTRACT

Three independent, but related studies were conducted to further the objective of improving soybean seed traits and resistance to tobacco ringspot virus. The objective of the first study was to identify and map the genes for bud blight caused by tobacco ringspot virus. Two populations derived from the cross of Young x PI416937 were evaluated. In the field, the plots were naturally infected with tobacco ringspot virus and were visually scored. RFLP and SSR markers were used to map the gene for bud blight. Composite interval mapping identified a major QTL for bud blight resistance between the RFLP markers K644_1 and A069_b. The major QTL was also linked to the Satt510 marker.

The objective of the second study was to utilize an independent F_2-derived soybean population of PI97100 x Coker 237 to confirm previously reported RFLP markers associated with seed protein, seed oil, and seed weight, mapped in an F_2-derived population created from the same parents. Single-factor analysis of variance was used to confirm the RFLP loci based on the mean phenotypic data across three different environments. Two out of four previously reported QTL for seed protein, two out of three QTL for seed oil content, and one out of three QTL for seed weight were confirmed in the independent population.

The objective of the third study was to investigate the presence of genetic variation for seed composition traits within three elite soybean cultivars by honeycomb selection. Divergent honeycomb selection for seed protein and oil within each cultivar was performed using a replicated-3 honeycomb design. The selected plants were evaluated in replicated row-plot experiments for three years. Honeycomb selection was successful in discovering a significant amount of within-cultivar genetic variation for seed protein and oil. Across the three cultivars, the magnitude of within-cultivar variation averaged 5% (19 g/kg) for protein and 6% (12 g/kg) for oil content. The magnitude of within-cultivar variation for fatty acid content ranged from 6 to 29 g/kg across the three cultivars.

INDEX WORDS: Soybean, Quantitative trait loci, QTL, Genetic mapping, RFLP, SSR, Bud blight, Tobacco ringspot virus, Seed protein, Seed oil, Seed weight, Honeycomb selection, Honeycomb breeding, Genetic variation, Within-cultivar variation.

SOYBEAN IMPROVEMENT OF SEED TRAITS AND RESISTANCE TO BUD BLIGHT VIA MOLECULAR MARKERS AND CLASSICAL METHODS

by

VASILIA APOSTOLOS FASOULA

FORB, Aristotle University of Thessaloniki, Greece, 1990

M.S., University of Illinois at Urbana-Champaign, 1994

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2002

SOYBEAN IMPROVEMENT OF SEED TRAITS AND RESISTANCE TO BUD BLIGHT VIA MOLECULAR MARKERS AND CLASSICAL METHODS

by

VASILIA APOSTOLOS FASOULA

Approved:

Major Professor: H. Roger Boerma

Committee: Joe Bouton
John All
Jerry Johnson
Paul Raymer

Electronic Version Approved:

Gordhan L. Patel
Dean of the Graduate School
The University of Georgia
August 2002

DEDICATION

To my loving husband Dimitri

and our wonderful parents Apostolos-Maria, Periklis-Nitsa

ACKNOWLEDGMENTS

I would like to thank the head of the department, Dr. Al Smith, and my committee members, Dr. Roger Boerma, Dr. Joe Bouton, Dr. John All, Dr. Jerry Johnson, and Dr. Paul Raymer for their constructive criticism and their understanding and willingness to accommodate me. I greatly enjoyed the scientific interaction with each one of them, and I feel very privileged and fortunate to have known them.

I would also like to express my gratitude to the following people: Dale Wood and Gina Rowan for their assistance with the field experiments and with Agrobase. Barbara Stewart, Jennie Alvernaz, David Walker, and David Hulbert for their assistance and advice in the laboratory. Carolyn Hightower, for being so helpful with administrative tasks and especially for her sweet personality. Several past and present graduate students and postdoctoral associates, for it has been my greatest pleasure to exchange ideas and interact with them. My dear friends Georgia, Frieda, and Panos, for their help and support.

Many special thanks go to; Dr. Glenn Burton and his family for the support they provided to me through the “Glenn and Helen Burton Feeding the Hungry Scholarship”. My mother, for her unconditional and endless love. My father, for being a great inspiration to me and for showing me that creativity and enthusiasm, critical thinking and open-mindedness, communication and teaching, kindness and understanding are invaluable skills that drive progress and promote breakthroughs in science. And my

loving husband Dimitri, for his continuous support, help, understanding, and love in so many ways.

Finally, I would like to thank my mentor and teacher, Roger Boerma, for his guidance, constructive criticism, consideration, and support throughout my graduate school years. The stimulating scientific discussions we had at the farm, while working on the field experiments, as well as in the office and laboratory, while elaborating on my research projects, will be kept dear in my heart forever.

TABLE OF CONTENTS

“No two plants are exactly alike. They’re all different, and as a consequence, you have to know that difference. I start with the seedling, and I don’t want to leave it. I don’t feel I really know the story if I don’t watch the plant all the way along. So I know every plant in the field. I know them intimately, and I find it a great pleasure to know them.”

“Genome complexity is such as to enable it not only to program the life cycle of the organism, with fidelity to past and future generations, but also to reprogram itself when exposed to sufficient environmental stress. Thereby, effecting a kind of learning from the organisms experience. Such a picture implies a concept of genetic variation that is neither random not purposive.”

Barbara McClintock

CHAPTER 1

INTRODUCTION

Soybean [*Glycine max* (L.) Merr.] belongs to the legume family and is native to Asia. Soybean seed is a major source of protein for animal feed and oil for human consumption. It supplies approximately 70% of the world's protein meal and 25% of the world's supply of vegetable oil. The high concentration of protein in soybean meal makes the meal a valuable livestock feed. Soybean protein has an excellent balance of amino acids compared with other vegetable proteins. Soybean protein is complete in that it contains all eight essential amino acids needed for human health. In the United States, the annual production of soybean meal is approximately 34,686 Mg (Golbitz, 2001).

Soybean improvement most commonly involves crosses between elite cultivars to generate novel recombinant genotypes from which superior cultivars are selected. Genotypes are advanced by self-pollination to the F_4 or F_5 generation where the genomes are considered nearly fixed. F_5-derived lines are then evaluated in replicated field trials and multiple locations to select the ones that exhibit superior agronomic performance and adaptability over a wide range of environments.

Genetic improvement in soybean for most quantitative traits is complex because of the number of genes involved in the expression of these traits and the interaction of these genes with the environment. With the advance of molecular genetics, new tools are

now available to plant breeders. A powerful approach for the improvement of agronomically important traits is the development of DNA marker linkage maps that have the potential to saturate the genome, thus increasing the possibility of finding significant associations between molecular markers and quantitative traits (Keim et al., 1990a). These linkages can then be utilized in marker-assisted selection.

With the U.S.A. being the leading producer of soybean in the world, breeding programs have to integrate resistance to pests and diseases into their soybean cultivars. Most cultivars released in the southern U.S. are resistant to southern stem canker, bacterial pustule, frogeye leaf spot, root-knot nematodes, and some prevalent races of soybean cyst nematode.

Soybean breeding programs normally restrict the use of parents to those improved for a variety of traits. Progress is still sufficient to encourage continued breeding within narrow gene pools although each cycle is expected to reduce genetic variation. A lot of interest exists in genetic diversity, primarily due to the belief that progress will be hampered without adequate genetic variation. On the contrary, evidence from selection experiments within fairly homogeneous genetic pools suggests that the genome is more flexible and plastic than previously assumed (Sprague et al., 1960; Russell et al., 1963; Fasoula, 1990; Dudley and Lambert, 1992; Rasmusson and Phillips, 1997; Fasoulas, 2000).

Elite cultivars possess a limited amount of genetic variation, which is generally believed to be too small to make significant progress. The release of cultivars is a time consuming task, thus it is important to exploit any positive source of latent or newly created genetic variation. Honeycomb breeding emphasizes selection after the release of

cultivars to further improve the crop yield and quality of elite cultivars and to circumvent cultivar deterioration (Fasoula and Fasoula, 2000). The most fundamental principle of honeycomb breeding is that selection and evaluation is performed on single plants grown in the absence of competition (Fasoula and Fasoula, 1997). Honeycomb designs are utilized to sample effectively for environmental diversity and to select objectively the most superior entries as well as the best plants within the selected entries (Fasoulas and Fasoula, 1995).

The objectives of this research were to: (i) identify, map, and confirm the gene(s) for bud blight caused by tobacco ringspot virus; (ii) confirm previously identified QTL for seed protein, seed oil content, and seed weight in a F_2-derived population of PI97100 x Coker 237 using RFLP markers; (iii) investigate the presence of genetic variation for seed protein and oil within three elite soybean cultivars by honeycomb selection.

CHAPTER 2

REVIEW OF LITERATURE

Introduction

Soybean [*Glycine max* (L.) Merrill] is an important food crop and ranks second behind corn (*Zea mays*) in total hectares grown for a seed crop in the U.S.A. During the past 25 years, U.S. soybean yields continue to increase 1.0 to 1.5% per year (American Soybean Association, 2001). It is estimated that at least 50% of this gain is solely due to the development of improved cultivars. Improvement most commonly involves crosses between elite cultivars to generate novel recombinant genotypes from which superior cultivars are selected. Genotypes are advanced by self-pollination to the F_4 or F_5 generation where the genomes are considered nearly fixed. F_5-derived lines are then evaluated in replicated field trials and multiple locations to select the ones that exhibit superior agronomic performance and adaptability over a wide range of environments.

Selection for agronomic traits such as plant height, lodging, and maturity has been extensively applied in soybean breeding programs for the development of cultivars with superior performance. These traits possess relatively high heritabilities (>50%) compared to seed yield (Anand and Torrie, 1963; Kwon and Torrie, 1964), but they also exhibit complex polygenic inheritance. Soybean seed is a major source of protein for animal feed

and oil for human consumption. It supplies approximately 70% of the world's protein meal and 25% of world's supply of vegetable oil. The high concentration of protein in soybean meal makes the meal a valuable livestock feed. Soybean protein has an excellent balance of amino acids compared with other vegetable proteins. In the United States, the annual production of soybean meal is approximately 34,686 Mg (Golbitz, 2001). More than 95% of this meal is used as animal feed for poultry, swine, and cattle.

Soybean oil is an important source of vegetable oil for human food and accounts for 22% of the world's total edible oil production (Glaudemans et al., 1998). However, crude soybean oil has flavor and stability limitations that can be overcome only by expensive processing steps. The flavor problem of soybean oil has been attributed to the oxidation of fatty acids, primarily linolenic acid (Smouse, 1979). Soybean oil consists mainly of palmitic (C16:0), stearic (C18:0), oleic (C18:1), linoleic (C18:2), and linolenic (C18:3) acid. New soybean genotypes with modified fatty acid composition are constantly being considered for commercial production. A reduction in saturated fatty acids in soybean oil is desirable in order to reduce the human health risks associated with the consumption of saturated fats. Simultaneous increases in protein and oil content can proceed only to a limited extent, and most experimental data show that protein and oil content are negatively correlated (Johnson and Bernard, 1962; Kwon and Torrie, 1964; Smith and Weber, 1968; Burton, 1987).

Substantial progress has been accomplished toward the selection of lines with higher percentage of protein and oil content in the seeds (Brim and Burton, 1979; Burton, 1985; Burton and Brim, 1981). Heritability estimates for percent protein range from 60% to 90% and those for percent oil from 50% to 90% (Byth et al., 1969; Fehr and Weber,

1968; Kwon and Torrie, 1964). Seed weight is a highly heritable trait with heritability estimates ranging from 44 to 94% (Brim, 1973). Seed weight is one of the yield components of soybean and in general is positively correlated with seed yield (Burris et al., 1973; Smith and Camper, 1975).

Utilization of molecular markers in soybean breeding

Genetic improvement in soybean for most quantitative traits is complex because of the number of genes involved in the expression of these traits and the interaction of these genes with the environment. With the advance of molecular genetics, new tools are now available to plant breeders. A powerful approach for the improvement of agronomically important quantitative traits is the development of DNA marker linkage maps that have the potential to saturate the genome, thus increasing the possibility of finding significant associations between molecular markers and quantitative traits (Keim et al., 1990a). These linkages can then be utilized in marker-assisted selection.

A variety of molecular markers are currently available. These include restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), microsatellites or simple sequence repeats (SSR), and amplified fragment length polymorphism (AFLP). RFLPs can detect variation produced by both base substitution and insertion/deletion and require the use of single copy DNA sequences. RAPDs utilize short random primers and the polymerase chain reaction (PCR) and are better for detecting base substitution (Staub et al., 1996). SSRs are specific for tandemly repeated DNA in the genome and usually only detect insertions/deletions. AFLP markers provide

great capacity for large scale mapping and are based on RFLPs that are detected after selective polymerase chain reaction (Staub et al., 1996; Powell et al., 1996).

Soybean RFLP (restriction fragment length polymorphism) markers were introduced in the late-1980s (Apuya et al., 1988; Keim et al., 1989). The arbitrarily primed marker systems (RAPDs) were developed in the early 1990's (Caetano-Anolles et al., 1991; Williams et al., 1990). Soybean SSR (short sequence repeat) markers were introduced in 1992 (Akkaya et el., 1992). In the past decade, several high density genetic maps have become available (Shoemaker et al., 1996; Keim et al., 1997; Cregan et al., 1999). The classical genetic map of soybean has evolved over the course of the last half century (Shoemaker and Specht, 1995). The integrated genetic linkage map of the soybean genome consists of RFLP, RAPD, SSR, AFLP, isozyme, and classical markers (Cregan et al., 1999).

The concept of using genetic markers is based on the fact that the marker locus serves to identify or mark a specific chromosomal region and enables that linked region to be followed through various genetic manipulations (Stuber et al., 1989). In principle, such dissection of polygenic traits is straightforward: i) a backcross or intercross is performed between two lines differing in a trait of interest; ii) progeny are scored both for the trait and for markers spaced throughout the genome; and iii) a correlation is sought between the trait and the inheritance pattern of one or more markers (Lander and Botstein, 1989).

A number of methods of identifying associations between marker alleles and quantitative traits have been proposed (Dudley, 1993). The most widely used are linear regression, interval mapping, and composite interval mapping. Linear regression involves

comparisons among the phenotypic means of appropriate marker classes of progeny. The significance of the association is assessed by analysis of variance and evaluated by the appropriate *F*-test. Phenotypic variation explained by markers is described by using the coefficient of variation (R^2), which is the proportion of the total variance among the phenotypic means explained by the marker-genotype classes.

Lander and Botstein (1989) proposed the use of interval mapping using a maximum likelihood approach. This method generalized the traditional single-marker analysis to the situation in which the QTL does not lie exactly at the marker locus but rather between two flanking markers. The strength of evidence of linkage is reflected in a LOD score, or logarithm to the base 10 of the likelihood ratio (Stuber et al., 1992). Essentially, the odds ratio denotes how much more probable it is for the data to have arisen if there is a QTL at the given position than if there is no linked QTL. The evidence for the presence of a QTL can be conveniently displayed by a QTL likelihood map, indicating the LOD score at all points along the length of a chromosome. When the LOD score exceeds a predetermined threshold, the presence of a QTL is declared (Lander and Botstein, 1989). Composite interval mapping combines the interval mapping with multiple regression (Zeng, 1994). Multiple regression analysis is useful for determining the most important markers among and within linkage groups and resolving whether multiple peaks are caused by single or multiple QTL on the same linkage group.

The use of RFLP markers or more recently other types of molecular markers has increased the efficiency of mapping quantitative trait loci (QTL) because of the greater number of markers that can be scored in a single population relative to other markers used such as isozyme or morphological markers (Diers et al., 1992b). RFLP analysis has

extensively been used in soybean in an effort to map various qualitative and quantitative loci. Some of the initial studies are cited below. Keim et al. (1990a) working with hard seededness in soybean, found five independent RFLP markers that explained 71 % of the variation in this trait. Further mapping (Keim et al., 1990b) revealed marker association with traits such as stem diameter, leaf length, leaf width, canopy height, and maturity. The R^2 values ranged between 16 and 24% for the independent markers.

Using a F_2 population, Landau-Ellis et al. (1991) and Landau-Ellis and Gresshoff (1992) detected one RFLP marker that was tightly linked to the locus that conditions supernodulation. Diers et al. (1992c) identified markers that were tightly linked to genes for resistance to *phytophthora* root rot. In another study, Diers et al. (1992b) used a F_2 population consisting of 60 plants and found three independent markers that were associated with significant variation for seed protein content. In addition, these QTL also conditioned seed oil content. Mansur et al. (1993) detected RFLP-QTL associations for maturity, plant height, lodging, and seed yield in a segregating population obtained from crossing the soybean cultivars Minsoy and Noir 1. Yu et al. (1994) found one SSR and two RFLP markers closely linked to the soybean mosaic virus resistance gene.

The utilization of RFLP markers for marker-assisted selection (MAS) in a breeding program is difficult because of their low polymorphic content and their high technical demand. Although RFLP-based maps provide a framework for MAS, mapping QTL with SSR markers would greatly facilitate MAS, because SSRs are abundant, highly polymorphic with minimal evidence of clustering, and practical for high throughput application (Akkaya et al., 1995; Maughan et al., 1995; Powell et al., 1996; Diwan and Cregan, 1997). In the past five years, the number of studies that reported QTL for the

various quantitative traits has grown exponentially. Boerma (2000) reviewed the information of mapped soybean QTL reported in literature and found that there are more than 300 QTL reported for the various quantitative traits. Of these 300 independent QTL, approximately 150 QTL were reported to condition 10% or more of the phenotypic variation in the trait.

GxE interactions for molecular markers and QTL

Precise estimation of the genotype by environment (GxE) interaction is essential in the assessment of mechanisms of inheritance as well as the prediction of performance in breeding programs because genotypic values must be inferred from phenotypic response (Stuber et al., 1992). Clearly, phenotypic performance depends on both genetic and nongenetic influences on plant growth and development. Consequently, the relative rankings of genotypes may differ in different environments and the relationship may be quite complex (Allard and Bradshaw, 1964).

A number of studies have attempted to discern the degree of GxE interaction at individual QTL by comparing the frequency of identification of significant DNA marker-QTL associations in different environments. In maize, Stuber at al. (1992) found that marker locus by environment interaction was relatively unimportant. Schön et al. (1994) also found relatively consistent QTL detection for morphological traits across environments in European maize lines grown in adapted environments. Veldboom and Lee (1996) reported that 50% of the QTL that were detected across environments were also detected in a stress and a nonstress environment. On the other hand, Bubeck et al.

(1993) and Paterson et al. (1991) found little agreement between environments in the markers identified as having significant association with quantitative traits.

In soybean, most studies that dealt with the detection of RFLP markers linked to important agronomic traits were conducted in a single environment (Keim et al., 1990a; Lark et al., 1993; Diers et al., 1992b, 1992c; Mansur et al., 1993). Lee et al. (1996b) tested RFLP-QTL associations over four locations in a population derived from the cross of Young x PI416937. Significant associations were detected for height, lodging, and maturity. Only two markers for plant height and one for lodging were detected in all locations, whereas good agreement of QTL across locations was found for maturity. These results indicate that the level of consistency of QTL across environments is trait specific (Lee et al., 1996b).

In another soybean population of a cross between PI97100 and Coker 237, Lee et al. (1996c) found that significant RFLP-QTL associations for plant height, lodging, and maturity were consistent across the two locations that were tested. This consistency was attributed to the fact that the identified markers are linked to genetic factors having major effects on the evaluated traits. Working with the same soybean populations, Bailey et al. (1997) and Mian et al. (1996) have found significant QTL associations across environments for pod dehiscence and seed weight, respectively. In another study, Lee et al. (1996d) identified QTL for seed protein and oil content that were consistent across locations. Brummer et al. (1997) mapped QTL for protein and oil content using eight soybean populations across a wide range of environments. For both protein and oil content, he found environmentally stable QTL that were detectable in most years and environmentally sensitive QTL that were detected only sporadically.

Confirming QTL in independent populations or generations

Although there are more than 300 QTL reported for the various quantitative traits in soybean, there is limited research on confirmation of the reported QTL and the results from these studies have been mixed. For example, Diers et al. (1992a) reported that none of the identified QTL conditioning iron deficiency chlorosis in soybean were effective for divergent selection among lines that were not used in the original QTL mapping study. Mudge et al. (1997) found that a single SSR (simple sequence repeat) marker was 95% accurate in predicting resistance to soybean cyst nematode. Moreover, when two SSR markers that flanked the QTL were used, the accuracy of predicting the resistant phenotype was increased to 98%. In another study, Li et al. (2001) reported that marker assisted selection at two QTL conditioning resistance to root-knot nematode was successful in identifying resistant lines in an independently derived soybean population.

Brummer et al. (1997) identified QTL for soybean seed protein and oil content using eight distinct populations. They reported that some QTL were sensitive to the environment that was tested and they detected both environmentally sensitive and stable QTL. Harris et al. (2001) utilized SSR markers in soybean in order to verify previously reported QTL associated with seed protein, seed oil, and seed weight in a newly created F_2-derived soybean population. In their study, SSR markers were selected from the consensus soybean map in the same genomic region as the RFLP markers identified in the original mapping study. They were able to confirm only 33% of the QTL reported in the original mapping study.

In maize (Zea mays L.), results from three independent experiments repeated in the same genetic background revealed that the QTL identified were not consistent (Beavis et al., 1994; Beavis, 1994). Confounding factors such as type of progeny, sources of parental lines, different sets of environments, and sampling of progeny have been reported as possible causes for this discrepancy (Beavis, 1994). In another maize study, Ajmone-Marsan et al. (1996) evaluated previously identified QTL for grain yield in an independent sample drawn from the same population. They found that two QTL were consistent with those detected in the previous experiments, but two QTL identified in the first sample remained undetected in the independent sample.

Melchinger et al. (1998) genotyped two independent samples of the same F_2 maize population using RFLP markers. QTL for grain yield and other agronomically important traits were mapped in both samples. For all traits, they detected a total of 107 QTL from the first sample and 39 QTL from the second independent sample. They found that only 20 QTL were common in both population samples. In another maize study, a comparison of QTL consistency across populations and generations based on common chromosomal positions revealed a low consistency of QTL positions across different populations and a moderate consistency between QTL found in different generations for a given population (Groh et al., 1998).

Most qualitative genetic studies in soybean require a hypothesis generation and a second or confirming generation in order to assign a gene symbol. However, this confirmation step has not been required in QTL mapping studies in soybean or other species (Boerma and Mian, 1999). Since there is limited and imprecise information confirming the reported QTL in soybean, it is important to conduct confirmation

experiments prior to the development of breeding strategies based on unconfirmed QTL reported in the literature. Furthermore, a lot of QTL mapping studies have not used multiple environments or populations for the collection of phenotypic data. Studies conducted in a single environment are likely to underestimate the number of QTL, which can influence a trait (Paterson et al., 1991).

The importance of verifying significant markers in an independent sample or population is of outmost value for plant breeders. Many researchers assume that once markers associated with a trait have been found, these will always explain a similar amount of variation in most environments. Plant breeders, on the other hand, recognize that the phenotype of a plant is determined by the interaction of its genotype with the environment and therefore phenotypic evaluations are dependent on the environmental conditions. Moreover, similar phenotypes can be a result of a different combination of genes expressed, meaning that marker-QTL associations may not always be the same when populations of the same or different crosses are evaluated.

Bud blight resistance

Bud blight, caused by tobacco ringspot virus (TRSV), can significantly reduce the seed yield and seed quality of soybean. TRSV is a member of the nepovirus group of plant viruses and consists of two single-stranded RNA molecules. TRSV affects the foliage and may cause almost total seed loss due to flower abortion. The virus moves into the roots and root nodules and significantly impairs symbiotic nitrogen fixation (Orellana, 1981). Symptoms of the virus include stunted growth, discoloration of stems and

branches, and curvature of the terminal meristem (Tu, 1986). The most striking symptom is the curvature of the terminal bud to form a crook. The buds on the plant become brown, necrotic, and brittle. TRSV infected buds may fall off at the slightest touch. Adventitious leaf and floral buds may proliferate excessively (Demski and Kuhn, 1999). Pods are generally severely underdeveloped or aborted and those that set before infection often develop dark blotches. These pods generally do not produce viable seeds and drop early. Root and nodule growth are also significantly reduced (Demski and Kuhn, 1999).

The mode of transmission of the virus in nature is not known, but the virus has been shown to be mechanically transmitted by grasshoppers (i.e., *Melanoplus differentialis* Thomas), nematodes (i.e., *Xiphinema americanum* Cobb), and thrips (*Thrips tabaci* Lind.) (Dunleavy, 1957; Bergeson et al., 1964). In soybean, the virus is transmitted primarily by infected seed and nematodes (Bergeson et al., 1964). TRSV infection is most detrimental when plants are infected during the early stages of plant development (Demski and Kuhn, 1999). Bud blight can cause 25 to 100% yield reduction in soybean (Crittenden et al., 1966). In China, bud blight is one of the three most important diseases of soybean and is considered very destructive in tropical and subtropical regions (Orellana, 1981).

Flor (1947) has defined the gene-for-gene model of host-pathogen interactions. According to this model, the pathogen produces an avirulence (*Avr*) factor that elicits resistance if the host carries a resistance (*R*) gene of particular specificity. In *Arabidopsis thaliana*, genetic studies revealed the identification of a single incompletely dominant locus (TTR1) controlling resistance to TRSV (Lee et al., 1996a). In another viral genetic study, Yu et al. (1994) used RFLP and SSR markers to identify and map the

chromosomal location of the soybean mosaic virus (SMV) resistance gene. They found that two RFLP markers, A186 and K644_1 were closely linked to the SMV resistance gene with distances of 1.5 and 2.1 cM, respectively. In addition, the RFLP markers A186 and K644_1 are linked with a Phytophthora root rot resistance gene in soybean (Diers et al., 1992c). Other studies reported that linkage relationships exist between the SMV resistance gene and the genes conditioning resistance to peanut mottle virus (PMV) and peanut stripe virus (PStV) (Roane et al., 1983; Choi et al., 1989). It appears that this genomic region in soybean encompasses a number of virus resistance genes and could be readily utilized in a marker-assisted breeding program.

Basic principles of the honeycomb breeding

The primary objective of the plant breeder is to identify and select superior genotypes from among and within a broad array of germplasm. Plant breeders attempt to select genotypes that perform well and are stable across a wide range of environments, including stress and non-stress environments. To accomplish this goal, the selection design should be capable of sampling effectively for spatial heterogeneity and ensuring comparable growing conditions for a large number of entries as well as for plants within entries. This allows efficient exploitation of soil heterogeneity and successful selection of genotypes characterized by high and stable yield.

Honeycomb designs (Fasoulas, 1993; Fasoulas and Fasoula, 1995) are systematic and possess two novel properties. The first is the allocation of entries across the experimental site in the corners of equilateral triangles. Figures 1 and 2 illustrate the

allocation of entries 21 and 10 across the experimental site, in the characteristic equilateral triangular lattice (ETL) pattern, which applies to all the evaluated entries and all the honeycomb designs. The ETL patterns ensure comparable growing conditions for all entries and effective sampling of soil heterogeneity (Fasoulas and Fasoula, 1995). Honeycomb designs are capable of evaluating a large number of entries (3-200 or more) without compromising experimental efficiency. Because of the nature of the designs, soil heterogeneity does not impose any restriction on the number of evaluated entries.

The second novel property of the designs is the formation of moving blocks or moving rings (Fig. 1). This property allows reliable selection of superior plants within the selected entries. For example, plants of entry 21 are always encircled by plants that belong to the same entries, thus a common multiplant check may be used. The same idea applies to all the entries (1 through 21) of the design (Fig. 1). Consequently, plants of a certain entry, regardless of their position in the field, are ranked objectively since their yield is expressed as a percentage of the average yield of a common multiplant denominator that may include any number of plants (Fasoula and Fasoula, 2000). The number of the plants within a moving ring represents the chosen selection pressure. For example, the first moving ring consists of 7 plants, the second moving ring consists of 13 plants, the third one consists of 19 plants, and so on (Fig. 2). Moving-ring selection is applied to every plant in the field in order to select the best plants within the most superior entries.

Honeycomb breeding is based on six fundamental principles: (i) selection and evaluation of single plants in the absence of competition. The masking effect of competition on the efficiency of single-plant selection has been reviewed by Fasoula and

Fasoula (1997a); (ii) selection and evaluation of plants under enhanced gene fixation, to fully exploit additive genetic variation that is responsible for progress through selection (Fasoula and Fasoula, 1997b); (iii) early and consistent multiple environment screening; (iv) utilization of honeycomb designs; (v) analysis of crop yield into three genetic components that can be assessed effectively on the single plant level; and (vi) continuous selection for crop yield and quality of elite cultivars in order to maintain and/or improve their productivity and quality over the years.

Partition of crop yield into genetic components

Honeycomb breeding has been evolving over the years. An important new addition in the methodology is the partition of crop yield into three genetic components and the development of criteria that correlate crop yield performance with single-plant yield performance. Crop yield potential has been partitioned into the following genetic components: (i) yield per plant, (ii) tolerance to stresses, and (iii) responsiveness to inputs (Fasoula and Fasoula, 2000). The first component involves genes controlling yield potential per plant and is estimated by the progeny mean. The second component involves genes conferring tolerance to the biotic and abiotic stresses, and is estimated by the progeny standardized mean. The third component involves genes for responsiveness to inputs and is estimated by the progeny standardized selection differential. The progeny mean, the progeny standardized mean, and the progeny standardized selection differential are calculated from the values of single plants representative of every progeny line grown in honeycomb experiments. Effective selection for genes controlling the three

components of crop yield leads to the development of density-independent cultivars with high and stable productivity and quality.

Density-independent cultivars yield optimally over a wide range of plant densities and are favored by the farmers. For example, Tokatlidis et al. (2001) showed that joint selection in the absence of competition for high yield per plant and tolerance to stresses accelerates the development of density-independent maize hybrids. Improved inbred lines were extracted from the F_2 generation of the single-cross hybrid PR-3183 by applying honeycomb selection, based on progeny mean yield per plant and progeny stability as assessed by the progeny standardized mean (Tokatlidis et al., 1998; 1999). Twelve inbred lines were chosen to develop six hybrids that were in essence "recycled" since they were derived from PR-3183. The recycled hybrids were found to be less density-dependent than the original hybrid PR-3183 because they had improved performance at the low plant density of 2.5 plants/m^2 and equal performance at the high plant density of 8.3 plants/m^2 (Tokatlidis et al., 2001).

Previous research on honeycomb selection

Mitchell et al. (1982) applied honeycomb selection in durum wheat (*Triticum turgidum* L.) to compare selection response for yield between 30- and 60-cm plant spacings (12.8 plants/m^2 and 3.2 plants/m^2, respectively). They concluded that selection for single-plant yield was more effective at wider plant spacings. In spring wheat, Lungu et al. (1987) assessed the merits of divergent honeycomb selection for yield in the F_2 and F_3 generations on plants spaced 50 cm apart (4.6 plants/m^2). He obtained two

significantly different populations whose performance carried over when tested at traditional crop densities. They concluded that honeycomb selection for yield could be used effectively for early generation selection in spring wheat.

Robertson and Frey (1987) used a honeycomb arrangement with oat (*Avena sativa* L.) plants spaced 45 cm apart (5.7 plants/m^2) and concluded that selection for grain yield in a noncompetitive field arrangement identified high-yielding oat lines. They found that there was a positive relationship between single-plant and line mean yields, which resulted in a 9.8% gain in grain yield. In the studies above, the interplant spacings used were not ideal because they still involved competition. In cereals, the ideal interplant spacing that does not involve competition can be found between 1.8 plants/m^2 and 1.2 plants/m^2.

Kyriakou and Fasoulas (1985) applied three intensities of honeycomb selection, (14.3, 5.3 and 1.6%) to a rye (*Secale cereale* L.) population under competition (51.3 plants/m^2) and nil-competition (1.4 plants/m^2) environments. The selected materials and the original population were grown in replicated trails at commercial crop densities. In the materials selected under competition, response to selection decreased when a greater selection pressure was applied. On the contrary, in the materials selected under nil-competition, response to selection increased when a larger selection pressure was applied.

Xanthopoulos (1990) applied honeycomb selection within an open-pollinated sunflower (*Helianthus annuus*) cultivar under competition (5.7 plants/m^2) and nil competition (0.7 plants/m^2). Only seven plants from each population were selected. When the two populations were evaluated at crop densities, the population selected at nil competition gave a 26% response to selection and a 66% realized heritability, whereas

the population selected under competition gave a 8.2% response to selection and a 30% realized heritability.

Kulkarni (1991) applied three cycles of honeycomb selection in the aromatic plant davana (*Artemisia pallens* Wall.) and found that the effects of selection performed on widely spaced plants were carried over when the progenies were grown at high plant density. He obtained a herb yield response of 12.4% per cycle without having any adverse effect on essential oil contend, davanone content in oil, or plant height. Gill et al. (1995) evaluated the comparative efficiency of four selection methods in mungbean [*Vigna radiata* (L.) Wilczek]. They concluded that the honeycomb method showed a marked superiority for yield over the pedigree selection, the single-seed descent, and the bulk method and was more efficient in isolating superior lines with high seed yield and pod number.

Batzios et al. (2001) evaluated the efficiency of honeycomb and conventional pedigree methods of selection for yield and fiber quality in cotton (*Gossypium* spp.). They reported that in both populations studied, honeycomb selection was more effective than conventional pedigree selection in identifying lines with high yielding ability and good lint quality. Ntanos and Roupakias (2001) evaluated the efficiency of honeycomb selection and panicle-to-row selection in two F_2 rice (*Oryza sativa* L.) populations. They reported that the average superiority of the honeycomb method over the panicle-to-row method in the two populations was 6 and 5% for grain yield, 18 and 9% for grain vitreosity, and 3 and 2% for grain length/width ratio. Additional data on the honeycomb breeding have been reviewed by Fasoula and Fasoula (2000).

Selection within elite gene pools

As the world's major crops continue under intense selection, interest in genetic diversity is growing and has been described in many crops including major field crops such as maize (*Zea mays* L.), soybean, and wheat (*Triticum aestivum* L.). Breeding programs in major crops normally restrict the use of parents to those improved for a variety of traits. Progress is still sufficient to encourage continued breeding within narrow gene pools although each cycle is expected to reduce genetic variation. There is a lot of interest in genetic diversity, primarily due to the belief that progress will be hampered without adequate genetic variation. On the contrary, evidence from selection experiments within fairly homozygous genetic pools suggests that the genome is more flexible and plastic than previously assumed. In her Nobel address, McClintock (1984) argued that the genome is dynamic and that it could modify itself under a variety of stress conditions.

One of the first researchers to report the existence to genetic variability within elite gene pools was Manning (1955). He worked with the cotton cultivar BP52, which was derived from a single-plant selection of the cultivar Nyasaland Upland. This selection attained cultivar status through bulking of open pollinated seed. He found an appreciable amount of genetic variation within BP52 and was able to make significant yield improvements that were not expected based on the theoretical estimates of genetic variances for strains and progenies within strains. One of his explanations for the persistence of genetic variability in inbred cotton was fortuitous crossovers in chromosome segments that released variation favorable to selection.

Gordon and Byth (1972) reported significant variation within the predominantly self-pollinated tobacco (Nicotiana tabacum L.) cultivar Hicks. The authors found large differences for several agronomic traits among selections derived from Hicks and suggested that substantial advances can be made from selection among these strains. Byth and Weber (1968) found genetic variability within F_5-derived soybean lines that are considered to be relatively homozygous.

Evidence from long-term selection experiments and doubled haploid studies also suggest that the genome is more flexible and plastic than previously assumed. The Univ. of Illinois long-term selection studies for modified oil and protein in maize are especially interesting (Dudley and Lambert, 1992). These lines have been selected for more than 90 generations and variation is still enough to achieve progress from selection. Sprague et al. (1960) and Russell et al. (1963) reported that doubled haploid lines of maize soon accumulated considerable variation in agronomic traits. These lines should have been homozygous at every locus and the authors argued that the observed variation could not be accounted for by the commonly reported rates of mutation.

Furthermore, Rasmusson and Phillips (1997) reported that in barley (*Hordeum vulgare* L.) incremental genetic gains were made for several traits in what appears to be a very narrow gene pool. They emphasized that elite gene pools have inherent mechanisms to provide a continuing source of new genetic variation. They hypothesized that selection gain occurs due to variation present in the original gene pool as well as due to de novo generated variation. Classical and molecular genetic analyses have shown that many mechanisms exist to generate de novo variation, such as intragenic recombination, unequal crossing over, DNA methylation, transposable elements, and gene amplification

(Rasmusson and Phillips, 1997). A comparative RFLP analysis of sunflower cultivars (Zhang et al., 1995) and an evaluation of cultivar variation in rice using microsatellite and RFLP markers (Olufowote et al., 1997) revealed the existence of within-cultivar variation.

Honeycomb breeding emphasizes selection after the release of cultivars to further improve the crop yield and quality of elite cultivars and to circumvent cultivar deterioration. Fasoula (1990) worked with a long-established wheat cultivar Siete Cerros released by CIMMYT in 1966 and applied divergent honeycomb selection for yield in a population of 2,700 plants. Successful divergent selection within the wheat cultivar produced lines with 8% higher and 9% lower yields than Siete Cerros when evaluated at commercial crop densities. These data demonstrated that propagation of the wheat cultivar at crop densities, year after year, causes a progressive increase of the low yielding-strong competitor genotypes at the expense of the high yielding-weak competitor genotypes, which led to the gradual deterioration of Siete Cerros.

In cotton (*Gossypium hirsutum* L.), honeycomb selection for high yield per plant and quality was applied within 10,000 plants of the elite cultivar Sindos 80 and led to the development of the cultivar Macedonia. This cultivar exhibited a 10% average superiority over Sindos 80 across sixteen locations and two years. Honeycomb selection within Macedonia continued for two more years and the best lines were grown in honeycomb trials at two diverse sites, a *Verticillium*-infected and a *Verticillium*-free field. Honeycomb selection for high yield per plant was applied to both sites and led to the isolation of two lines that significantly outyielded the best check at both environments. These lines were tolerant to *Verticillium* wilt, whereas the original cultivar Sindos 80 was

susceptible to *Verticillium* (Fasoulas, 2000). In tomato (*Lycopersicon esculentum* L.), exploitable genetic variation was uncovered in advanced generations after the point of achieving theoretical homozygosity (F_7 generation) (Christakis and Fasoulas, 2002).

The concept of continuous selection is important and merits extensive testing because the breeder will gain fundamental knowledge on some important issues: (i) whether heritable variation is constantly being created, and (ii) whether cultivars deteriorate with time because they are not constantly being improved for yield and other desirable traits. Improvement of adapted cultivars saves time, labor, and effort because it reduces the number of crosses and populations the breeder evaluates every year. In addition, when value-added or other single-gene traits (i.e., glyphosate tolerance or the *Bacillus thuringiensis* resistance gene) are incorporated in a cultivar, it is important that the cultivar maintains and/or improves its good productivity and quality over the years.

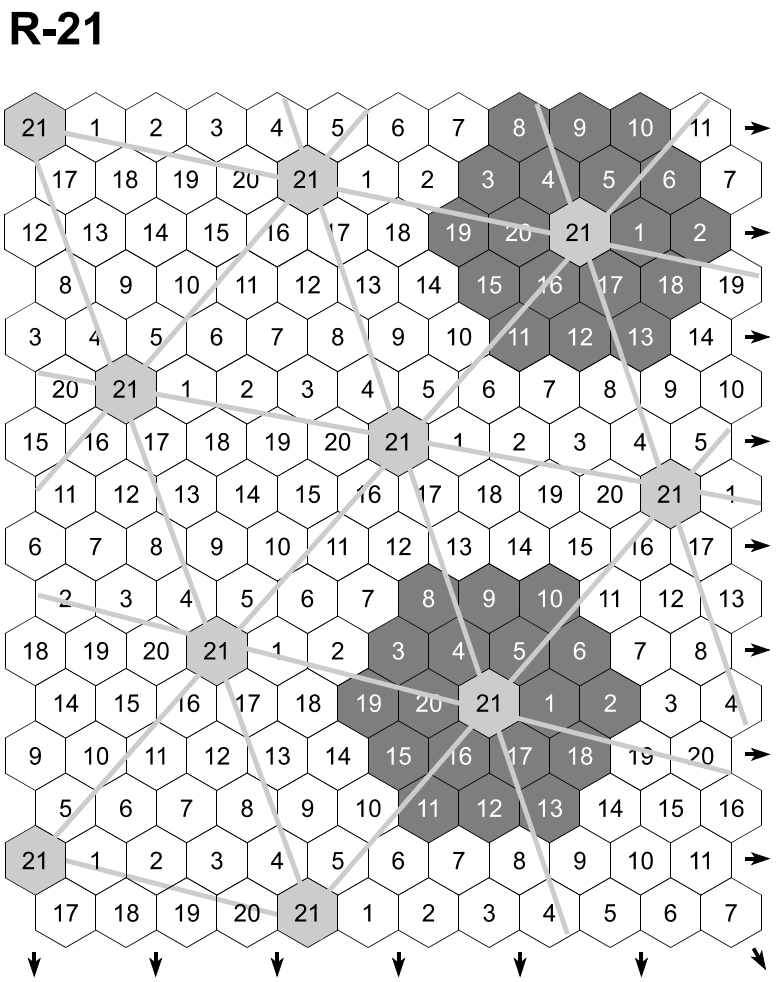

Figure 2.1. The replicated-21 honeycomb design evaluates a maximum of 21 entries. The numbers in the figure represent the position of the plants in the field and the hexagons represent the area exploited by each plant. Plants of any entry form an equilateral triangular lattice pattern (ETL) that samples effectively for spatial heterogeneity by allocating them evenly across the experimental site. The ETL pattern is exemplified for the plants of entry 21.

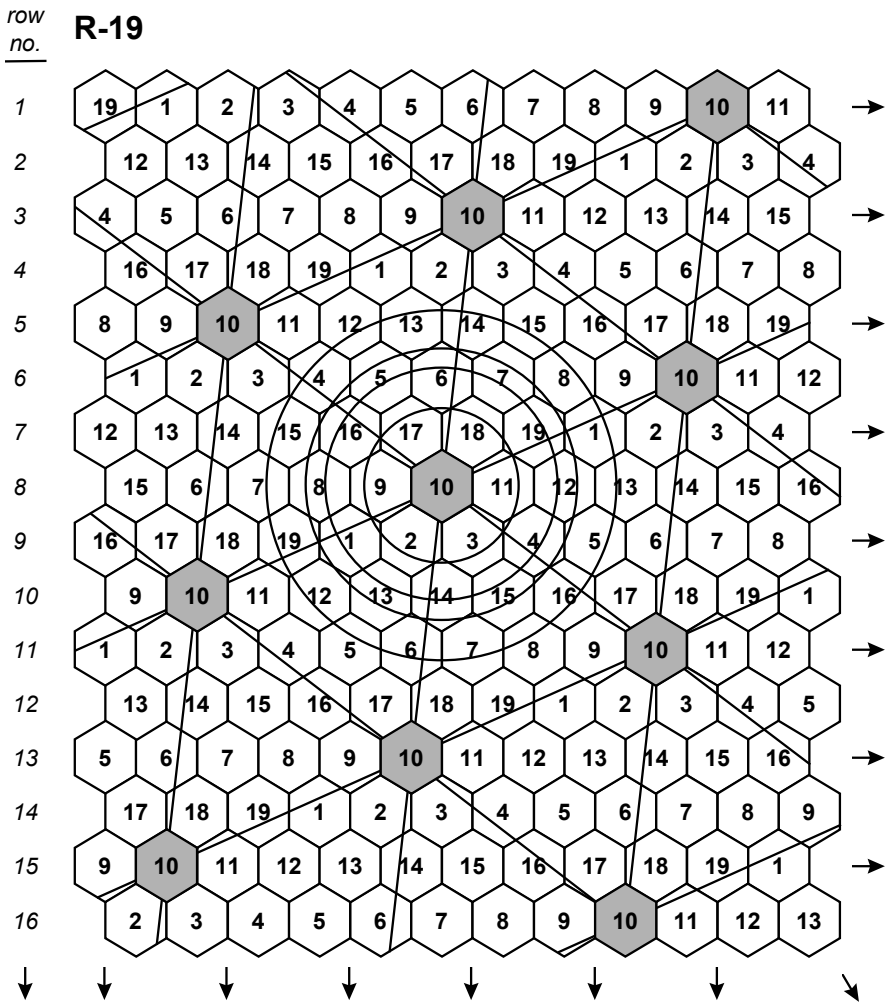

Figure 2.2. The replicated-19 honeycomb design evaluates plants from 19 entries, arranged in ascending order and in horizontal field rows. The ETL pattern is exemplified for the plants of entry 10. Four moving rings that represent four different selection pressures are illustrated for one plant of entry 10.

References

Ajmone-Marsan, P., G. Monfredini, A. Brandolini, A.E. Melchinger, G. Garay, M. Motto. 1996. Identification of QTL for grain yield in an elite hybrid of maize: repeatability of map position and effects in independent samples derived from the same population. Maydica 41:49-50.

Akkaya, M.S., A.A. Bhagwat, and P.B. Cregan. 1992. Length polymorphisms of simple sequence repeat DNA in soybean. Genetics 132:1131-1139.

Akkaya, M.S., R.C. Shoemaker, J.E. Specht, A.A. Bhagwat, and P.B. Cregan. 1995. Integration of simple sequence repeat DNA markers into a soybean linkage map. Crop Sci. 35:1439-1445.

American Soybean Association. 2001. 2001 Soy Stats. A reference guide to important soybean facts and figures. American soybean association, St. Louis, MO.

Allard, R.W., and A.D. Bradshaw. 1964. Implications of genotype-environmental interactions in applied plant breeding. Crop Sci. 4:503-508.

Anand, S.C. and J.H. Torrie. 1963. Heritability of yield and other traits and interrelationships among traits in the F_3 and F_4 generations of three soybean crosses. Crop Sci. 3:508-511.

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

Bailey, M.A., M.A.R. Mian, T.E. Carter, Jr., D.A. Ashley, and H.R. Boerma. 1997. Pod dehiscence of soybean: Identification of quantitative trait loci. J. Hered. 88:152-154.

Batzios, D.P., D.G. Roupakias, U. Kechagia, and S. Galanopoulou-Sendouca. 2001. Comparative efficiency of honeycomb and conventional pedigree methods of selection for yield and fiber quality in cotton (*Gossypium* spp.). Euphytica 122:203-211.

Beavis, W.D. 1994. The power and deceit of QTL experiments: lessons from comparative QTL studies. In: Proc of the Forty-nineth Annual Corn and Sorghum Industry Research Conference ASTA, Washington DC, pp 250-266.

Beavis, W.D., O.S. Smith, D. Grant, R. Fincher. 1994. Identification of quantitative trait loci using a small sample of topcrossed and F_4 progeny of maize. Crop Sci 34:882-896.

Bergeson, G.B., K.L. Athow, F.A. Laviolette, and Sister Mary Thomasine. 1964. Transmission, movement, and vector relationships of tobacco ringspot virus in soybean. Phytopathology 54:723-728.

Boerma, H.R. 2000. DNA markers: are they worth the cost? Proc. Am. Seed Trade Assoc. Soybean Seed Research Conf. Chicago Dec., 1999.

Boerma, H.R., and M.A.R. Mian. 1999. Soybean quantitative trait loci and marker-assisted breeding. World Soybean Res Conf VI Proc, Chicago, Illinois, pp 68-83.

Brim, C.A. 1973. Quantitative genetics and breeding. In: Caldwell B.E. (ed) Soybeans: improvement, production, and uses. American Society of Agronomy, pp. 155-186.

Brim, C.A., and J.W. Burton. 1979. Recurrent selection in soybeans. II. Selection for increased percent protein in seeds. Crop Sci. 19:494-498.

Brummer, E.C., G.L. Graef, J. Orf, J.R. Wilcox, and R. C. Shoemaker. 1997. Mapping QTL for seed protein and oil content in eight soybean populations. Crop Sci. 37:370-378.

Bubeck, D.M., M.M. Goodman, W.D. Beavis, and D.Grant. 1993. Quantitative trait loci controlling resistance to gray leaf spot in maize. Crop Sci. 33:838-847.

Burris, J.S., O.T. Edje, A.H. Wahab. 1973. Effect of seed size on seedling performance in soybeans. II. Seedling growth and photosynthesis and field performance. Crop Sci. 13:207-210.

Burton, J.W. 1985. Breeding soybeans for improved protein quantity and quality. In: Shibles R. (ed) Proc 3rd World Soybean Res. Conf., Westview Press, Boulder, Colorado, p. 361-367.

Burton, J.W. 1987. Quantitative genetics: Results relevant to soybean breeding. In: Wilcox, J.R. (ed) Soybeans: Improvement, production and uses. 2nd ed. Agron. Monogr. 16. ASA, CSSA, and SSSA, Madison, WI.

Burton, J.W., and C.A. Brim. 1981. Recurrent selection in soybeans. III. Selection for increased percent oil in seeds. Crop Sci. 21:31-34.

Burton, J.W., C.A. Brim, and M.F. Young. 1987. Registration of 'Young' soybean. Crop Sci. 27:1093.

Byth. D.E., and C.R. Weber. 1968. Effects of genetic heterogeneity within two soybean populations I. Variability within environments and stability across environments. Crop Sci. 8:44-47.

Byth, D.E., C.R. Weber, B.E. Caldwell. 1969. Correlated truncation selection for yield in soybeans. Crop Sci. 9:699-702.

Caetano-Anolles, G., B.J. Bassam, and P.M. Gresshoff. 1991. DNA amplification fingerprinting: a strategy for genome analysis. Plant Mol. Biol. Rep. 9:294-307.

Choi, S.H., S.K. Green, and D.R. Lee. 1989. Linkage relationship between two genes conferring resistance to peanut stripe virus and soybean mosaic. Euphytica 44:163-169.

Christakis, P.A., and A.C. Fasoulas. 2002. The effects of the genotype by environmental interaction on the fixation of heterosis in tomato. J. Agric. Sci. 138:1-5.

Cregan, P.B., T. Jarvik, A.L. Bush, K.G. Lark, R.C. Shoemaker, A.L. Kahler, T.T. vanToai, D.G. Lohnes, J. Chung, and J.E. Specht. 1999. An integrated genetic linkage map of the soybean. Crop Sci. 39:1464-1490.

Crittenden, H.W., K.M. Hastings, and D.M. Moore. 1966. Soybean losses caused by tobacco ringspot virus. Plant Dis. Reptr. 50:910-913.

Demski, J.W., and C.W. Kuhn. 1999. Tobacco ringspot. p. 66-68. In G.L. Hartman, J.B. Sinclair, and J.C. Rupe (ed.) Compendium of soybean diseases. Fourth edition. American Phytopathological Society, St. Paul, MN.

Diers, B.W., S.R. Cianzio, R.C. Shoemaker. 1992a. Possible identification of quantitative trait loci affecting iron efficiency in soybean. J Plant Nutrition 15:2127-2136.

Diers, B.W., P. Keim, W.R. Fehr, and R.C. Shoemaker. 1992b. RFLP analysis of soybean seed protein and oil content. Theor. Appl. Genet. 83:608-612.

Diers, B.W., L. Mansur, J. Ismande, and R.C. Shoemaker. 1992c. Mapping phytophthora loci in soybean with restriction fragment length polymorphism markers. Crop Sci. 32:377-383.

Diwan, N. and P.B. Cregan. 1997. Automated sizing of fluorescent-labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. Theor. Appl. Genet. 95:723-733.

Dudley, J.W. 1993. Molecular markers in plant improvement: Manipulation of genes affecting quantitative traits. Crop Sci. 33:660-668.

Dudley, J.W., and R.J. Lambert. 1992. Ninety generations of selection for oil and protein in maize. Maydica 37:81-87.

Dunleavy, J.M. 1957. The grasshopper as a vector of tobacco ringspot virus in soybean. Phytopathology 47:681-682.

Fasoula, D.A. 1990. Correlations between auto-, allo- and nil-competition and their implications in plant breeding. Euphytica 50:57-62.

Fasoula, D.A., and V.A. Fasoula. 1997a. Competitive ability and plant breeding. Plant Breed. Rev. 14:89-138.

Fasoula, D.A., and V.A. Fasoula. 1997b. Gene action and plant breeding. Plant Breed. Rev. 15:315-374.

Fasoula, V.A., and D.A. Fasoula. 2000. Honeycomb breeding: principles and applications. Plant Breed. Rev. 18:177-250.

Fasoulas, A.C. 1993. Principles of crop breeding. A.C. Fasoulas, P. O. Box 1555, Gr-54006 Thessaloniki 17, Greece.

Fasoulas, A.C. 2000. Building up resistance to *Verticillium* wilt in cotton through honeycomb breeding. p.120-124. In: F.M. Gillham (ed.) New frontiers in cotton research. Proc. 2nd World Cotton Research Conference. Sept. 6-12, 1998. Athens, Greece.

Fasoulas, A.C., and V.A. Fasoula. 1995. Honeycomb selection designs. Plant Breed. Rev. 13:87-139.

Fehr, W.R., and C.R. Weber. 1968. Mass selection by seed size and specific gravity in soybean populations. Crop Sci. 8:551-554.

Flor, H.H. 1947. Inheritance of reaction to rust in flax. J. Agric. Res. 74:241-262.

Gill, J.S., M.M. Verma, R.K. Gumber, and J.S. Brar. 1995. Comparative efficiency of four selection methods for deriving high-yielding lines in mungbean [*Vigna radiata* (L.) Wilczek]. Theor. Appl. Genet. 90:554-560.

Glaudemans, H.D., M.M.J. Timmermans, and H. Rijkse. 1998. The world of edible oils. p. 1-5. Rabobank International Marketing, Utrecht, the Netherlands.

Golbitz, P. 2001. 2001 Soya Bluebook Plus.

Gordon, I.L., and D.E. Byth. 1972. Comparisons among strains of the tobacco cultivar Hicks illustrating variability within a single cultivar. Queensland J. Agricultural and Animal Sci. 29:255-264.

Groh S., D. Gonzalez-de-Leon, M.M. Khairallah, C. Jiang, D. Bergvinson, M. Bohn, D.A. Hoisington, and A.E. Melchinger. 1998. QTL mapping in tropical maize. III. Genomic regions for resistance to *Diatraea* spp. and associated traits in two RIL populations. Crop Sci. 38:1062-1072.

Harris, D.K. 2001. Genetic improvement of soybean seed trait and resistance to bud blight and root-knot nematodes. PhD Thesis, Dept of Crop and Soil Sciences, Univ of Georgia.

Johnson, H.W., and R.L. Bernard. 1962. Soybean genetic and breeding. Adv. Agron. 14:149-221.

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

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

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

Keim, P., J.M. Schupp, S.E. Traivs, K.Clayton, T. Zhu, L. Shi, A. Ferreira, and D.M. Webb. 1997. A high-density soybean genetic map based on AFLP markers. Crop Sci. 37:537-543.

Kulkarni, R.N. 1991. Three cycles of honeycomb selection for herb yield in davana (*Artemisia pallens* Wall.). Euphytica 52:99-102.

Kwon, S.H. and J.H. Torrie. 1964. Heritability and interrelationships among traits of two soybean populations. Crop Sci. 4:196-198.

Kyriakou, D.T., and A.C. Fasoulas. 1985. Effects of competition and selection pressure on yield response in winter rye (*Secale cereale* L.). Euphytica 34:833-895.

Landau-Ellis, D., S.A. Ng, N. Angermuller, R.C. Shoemaker, and P.M. Gresshoff. 1991. The genetic locus controlling supernodulation in soybean (*Glycine max*) co-

segregates tightly with a cloned molecular marker. Mol. Gen. Genet. 228:221-226.

Landau-Ellis, D. and P.M. Gresshoff. 1992. Supernodulation soybean mutants nts382 and nts1007 map to the same RFLP marker confirming genetic complementation data. Soybean Genet. Newsl. 19:145-148.

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

Lark, K.G., J.M. Weisemann, B.F. Matthews, R. Palmer, K. Chase, and T. Macalma. 1993. A genetic map of soybean (*Glycine max* L.) using an intraspecific cross of two cultivars: Minsoy and Noir1. Theor. Appl. Genet. 86:901-906.

Lee, J-M., G.L. Hartman, L.L. Domier, and A.F. Bent. 1996a. Identification and map location of *TTR1*, a single locus in *Arabidopsis thaliana* that confers tolerance to tobacco ringspot nepovirus. Molecular Plant-Microbe Interactions 9:729-735.

Lee, S.H., M.A. Bailey, M.A.R. Mian, T.E. Carter, Jr., D.A. Ashley, R.S. Hussey, W.A. Parrott, and H.R. Boerma. 1996b. Molecular markers associated with soybean plant height, lodging, and maturity across locations. Crop Sci. 36:728-735.

Lee, S.H., M.A. Bailey, M.A.R. Mian, E.R. Shipe, D.A. Ashley, W.A. Parrott, R.S. Hussey, and H.R. Boerma. 1996c. Identification of quantitative trait loci for plant height, lodging, and maturity in a soybean population segregating for growth habit. Theor. Appl. Genet. 92:516-523.

Lee, S.H., M.A. Bailey, M.A.R. Mian, T.E. Carter, Jr., E.R. Shipe, D.A. Ashley, W.A. Parrott, R.S. Hussey, and H.R. Boerma. 1996d. RFLP loci associated with

soybean protein and oil content across populations and locations. Theor. Appl. Genet. 93:649-657.

Li, Z., L. Jakkula, R.S. Hussey, J.P. Tamulonis, H.R. Boerma. 2001. SSR mapping and confirmation of the QTL from PI96354 conditioning soybean resistance to southern root-knot nematode. Theor Appl Genet 103:1167-1173.

Lungu, D.M., P.J. Kaltsikes, and E.N. Larter. 1987. Honeycomb selection for yield in early generations of spring wheat. Euphytica 36:831-839.

Manning, H.L. 1955. Response to selection for yield in cotton. Cold Spring Harbor Symp. Quant. Biol. 20:103-110.

Mansur, L.M., K.G. Lark, H. Kross, and 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.

Maughan, P.J., M.A. Saghai-Maroof, and G.R. Buss. 1995. Microsatellite and amplified sequence length polymorphisms in cultivated and wild soybean. Genome 38:715-723.

McClintock, B. 1984. The significance of the responses of the genome to challenge. Science 226:792-801.

Melchinger, A.E., H.F. Utz, and C.C. Schön. 1998. Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. Genetics 149:383-403.

Mian, M.A.R., M.A. Bailey, J.P. Tamulonis, E.R. Shipe, T.E. Carter, Jr., W.A. Parrott, D.A. Ashley, R.S. Hussey, and H.R. Boerma. 1996. Molecular markers associated with seed weight in two soybean populations. Theor. Appl. Genet. 93:1011-1016.

Mitchell, J.W., R.H. Baker, and D.R. Knott. 1982. Evaluation of honeycomb selection for single plant yield in durum wheat. Crop Sci. 22-840-843.

Mudge, J., P.B. Cregan, J.P. Kenworthy, W.J. Kenworthy, J.H. Orf, N.D. Young. 1997. Two microsatellite markers that flank the major soybean cyst nematode resistance locus. Crop Sci 37:1611-1615.

Ntanos, D.A., and D.G. Roupakias. 2001. Comparative efficiency of two breeding methods for yield and quality in rice. Crop Sci. 41:345-350.

Olufowote, J.O., Y. Xu, X. Chen, W.D. Park, H.M. Beachell, R.H. Dilday, M. Goto, and S.R. McCouch. 1997. Comparative evaluation of within-cultivar variation of rice (*Oryza sativa* L.) using microsatellite and RFLP markers. Genome 40:370-378.

Orellana, R.G. 1981. Resistance to bud blight in introductions from the germ plasm of wild soybean. Plant Dis. 65:594-595.

Paterson, A.H., S. Damon, J.D. Hewitt, D. Zamir, H.D. Rabinowitch, S.E. Lincoln, E.S. Lander, S.D. Tanksley. 1991. Mendelian factors underlying quantitative traits in tomato: comparison across species, generations, and environments. Genetics 127:181-197.

Powell, W., M. Morgante, C. Andre, M. Hanafey, J. Vogel, S. Tingey, and A. Rafalski. 1996.The comparison of RFLP, RAPD, AFLP, and SSR (microsatellite) markers for germplasm analysis. Mol. Breeding 2:225-238.

Rasmusson, D.C., and R.L. Phillips. 1997. Plant breeding progress and genetic diversity from de novo variation and elevated epistasis. Crop Sci. 37:303-310.

Roane, C.W., S.A. Tolin, and G.R. Buss. 1983. Inheritance of reaction to two viruses in the cross York x Lee 68. J. Hered. 74:289-291.

Robertson, L.D., and K.J. Frey. 1987. Honeycomb selection among homozygous oat lines. Crop Sci. 27:1105-1108. Russell, W.A., G.F. Sprague, and L.H. Penny. 1963. Mutations affecting quantitative characters in long-time inbred lines of maize. Crop Sci. 3:175-178.

Russell, W.A., G.F. Sprague, and L.H. Penny. 1963. Mutations affecting quantitative characters in long-time inbred lines of maize. Crop Sci. 3:175-178.

Schön, C.C., A.E. Melchinger, J. Boppenmaier, E. Brunklaus-Jung, R.G. Herrmann, and J.F. Seitzer, 1994. RFLP mapping in maize: quantitative trait loci affecting testcross performance of elite European flint lines. Crop Sci. 34:378-389.

Shoemaker, R.C., K. Polzin, J. Labate, J.E. Specht, E.C. Brummer, T. Olson, N. Young, V. Concibido, J. Wilcox, J.P. Tamulonis, G. Kochert, and H.R. Boerma. 1996. Genome duplication in soybean (*Glycine* subgenus *Soja*). Genetics. 144:329-338.

Shoemaker, R.C., and J.E. Specht. 1995. Integration of the soybean molecular and classical genetic linkage maps. Crop Sci. 35:436-446.

Smith, T.J., and H.M. Camper, Jr. 1975. Effect of seed size on soybean performance. Agron. J. 67:681-684.

Smith, R.R., and C.R. Weber. 1968. Mass selection by specific gravity for protein and oil soybean populations. Crop Sci. 8:373-375.

Smouse, T.H. 1979. A review of soybean oil reversion flavor. J. Am. Oil Chem. Soc., 56:747A-751A.

Sprague, G.F., W.A. Russell, and L.H. Penny. 1960. Mutations affecting quantitative traits in selfed progeny of doubled monoploid maize stocks. Genetics 45:855-865.

Staub, J.E., F.C. Serquen, M. Gupta. 1996. Genetic markers, map construction, and their application in plant breeding. HortScience 31:729-741.

Stuber, C.W., S.E. Lincoln, D.W. Wolff, T. Helentjaris, and E.S. 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.

Tokatlidis, I.S., M. Koutsika-Sotiriou, A.C. Fasoulas, and A.S. Tsaftaris. 1998. Improving maize hybrids for potential yield per plant. Maydica 43:123-129.

Tokatlidis, I.S., M. Koutsika-Sotiriou, and A.C. Fasoulas. 1999. Load of deleterious genes and stability of performance in maize. Maydica 44:127-132.

Tokatlidis, I.S., M. Koutsika-Sotiriou, and A.C. Fasoulas. 2001. The development of density-independent hybrids in maize. Maydica 46:21-25.

Tu, J.C. 1986. Strains of tobacco ringspot virus isolated from soybean (*Glycine max*) in southwestern Ontario (Canada). Can. J. Plant Sci. 66:491-498.

Veldboom, L.R., and M. Lee. 1996. Genetic Mapping of quantitative trait loci in maize in stress and nonstress environments: I. Grain yield and yield components. Crop Sci. 36:1310-1319.

Williams, J.G.K., A.R. Kubelik, J.K. Livak, J.A. Rafalski, and S.V. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Mol. Gen. Genet. 244:638-645.

Xanthopoulos, F.P. 1990. Honeycomb mass selection efficiency for sunflower yield under different spacing conditions. Helia 13:121-125.

Yu, Y.G., M.A. Saghai Maroof, G.R. Buss, P.J. Maughan, and S.A. Tolin. 1994. RFLP and microsatellite mapping of a gene for soybean mosaic virus resistance. Phytopathology. 84:60-64.

Zhang, Y.X., L. Gentzbittel, F. Vear, and P. Nicolas. 1995. Assessment of inter- and intra-inbred line variability in sunflower (*Helianthus annuus*) by RFLPs. Genome 38:1040-1048.

Zeng, Z. 1994. Precision mapping of quantitative trait loci. Genetics 136:1457-1468.

CHAPTER 3

IDENTIFICATION, MAPPING, AND CONFIRMATION OF A GENE IN SOYBEAN FOR RESISTANCE TO BUD BLIGHT[1]

[1]Fasoula, V.A., D.K. Harris, M.A. Bailey, D.V. Phillips, and H.R. Boerma. Submitted to Crop Science.

Abstract

Bud blight, caused by tobacco ringspot virus (TRSV), can significantly reduce the seed yield and seed quality of soybean [*Glycine max* (L.) Merr.]. In this study, we evaluated two populations derived from the cross of 'Young' x PI416937. One population consisted of 116 F_4-derived lines and was used to identify and map restriction fragment length polymorphism (RFLP) markers associated with resistance to bud blight. Young is resistant to bud blight and PI416937 is susceptible. The lines were grown in one-row field plots in a randomized complete block design with two replications. The plots were naturally infected with tobacco ringspot virus. At the R2 stage of development, soybean plots were visually scored according to the number of plants that exhibited terminal bud death. A major QTL was identified and mapped by K644_1 on linkage group (LG) F. It accounted for 82% of the variation in bud blight score. Composite interval mapping run with QTL Cartographer located the single gene for bud blight resistance within the 8.2 cM interval between the RFLP markers K644_1 and A069_b. To verify the genomic location of the major bud blight QTL, a second population of Young x PI416937 that consisted of 180 F_2-derived lines was evaluated. In this population, simple sequence repeat (SSR) markers were utilized. The major QTL conditioning bud blight resistance was found to be closely linked to the Satt510 marker, which is approximately 2 to 4 cM from RFLP marker K644_1 based on the USDA/Iowa State Univ. soybean genetic map. Satt510 can be used successfully in marker-assisted selection for bud blight resistance to eliminate lines that are susceptible to bud blight.

Introduction

Bud blight, caused by tobacco ringspot virus (TRSV), can significantly reduce the seed yield and seed quality of soybean. TRSV is a member of the nepovirus group of plant viruses and consists of two single-stranded RNA molecules. TRSV affects the foliage and may cause almost total seed loss due to flower abortion. The virus moves into the roots and root nodules and significantly impairs symbiotic nitrogen fixation (Orellana, 1981). Symptoms of the virus include stunted growth, discoloration of stems and branches, and curvature of the terminal meristem (Tu, 1986). The most striking symptom is the curvature of the terminal bud to form a crook. The buds on the plant become brown, necrotic, and brittle. TRSV infected buds may drop from the plant at the slightest touch. Adventitious leaf and floral buds may proliferate excessively (Demski and Kuhn, 1999). Pods are generally severely underdeveloped or aborted and those that set before infection often develop dark blotches. These pods generally do not produce viable seeds and drop early. Root and nodule growth are also significantly reduced (Demski and Kuhn, 1999).

The mode of transmission of the virus in nature is not known, but the virus has been shown to be mechanically transmitted by grasshoppers (i.e., *Melanoplus differentialis* Thomas), nematodes (i.e., *Xiphinema americanum* Cobb), and thrips (*Thrips tabaci* Lind.) (Dunleavy, 1957; Bergeson et al., 1964). In soybean, the virus is transmitted primarily by infected seed and nematodes (Bergeson et al., 1964). TRSV infection is most detrimental when plants are infected during the early stages of plant development (Kahn et al., 1955; Demski and Kuhn, 1999). Bud blight can cause 25 to

100% yield reduction in soybean (Crittenden et al., 1966). In China, bud blight is one of the three most important diseases of soybean and is considered very destructive in tropical and subtropical regions (Orellana, 1981).

Flor (1947) has defined the gene-for-gene model of host-pathogen interactions. According to this model, the pathogen produces an avirulence (*Avr*) factor that elicits resistance if the host carries a resistance (*R*) gene of particular specificity. In *Arabidopsis thaliana*, genetic studies revealed the identification of a single incompletely dominant locus (*TTR1*) controlling resistance to TRSV (Lee et al., 1996a).

DNA marker technology has been readily developed and integrated into soybean breeding programs (Cregan et al., 1999; Boerma and Mian, 1999). Verification and confirmation of the QTL are recommended prior to the application of marker-assisted selection in a practical breeding program (Boerma and Mian, 1999). Simple sequence repeat (SSR) markers have recently been developed in soybean, and they are abundant, highly polymorphic, and distributed throughout the genome (Rongwen et al., 1995; Cregan et al., 1999). SSR markers are highly amenable for automation and allele sizing, which allows their high-throughput application and makes them an excellent source of DNA markers in a breeding program (Diwan and Cregan, 1997; Mitchell et al., 1997).

The first objective of this study was to utilize restriction fragment length polymorphism (RFLP) markers to identify and map the QTL that confers resistance to bud blight in the soybean population of Young x PI416937. The second objective was to confirm the genomic location of the bud blight QTL by utilizing SSR markers in an independent population of Young x PI416937.

Materials and Methods

Mapping population

A soybean population of 116 F_4-derived lines from the cross of Young x PI416937 was created by single-seed-descent with each line originating from a different F_2 plant (mapping population). Young is a highly productive cultivar that belongs to Maturity Group VI (Burton et al., 1987). DNA isolation, restriction enzyme digestion, electrophoresis, southern blotting, and hybridization procedures were performed according to Lee et al. (1996b). RFLP probes used were cDNA and genomic clones originated from *Glycine max*, *Vigna radiata* L., *Phaseolus vulgaris* L., *Arachis hybogaea* L., and *Medicago sativa* L. RFLP polymorphic probes with respect to the parents were used for mapping. A linkage map was constructed with marker data using the Kosambi map function of Gmendel 3.0, as if the data were derived from F_4 lines (Holloway and Knapp, 1993). For grouping linked markers, a minimum LOD (Likelihood of odds) of 3.0 and a maximum recombination frequency (rmax) of 0.38 (approximately equal to 50 cM) were used.

The parents and the 116 F_4-derived lines were planted on 25 May 1995 near Athens, GA at the Univ. of Georgia Plant Sciences Farm, in a randomized complete block design with two replications. The entries were grown in one-row plots with a row length of 3 m and a row spacing of 97 cm. The plots were naturally infected with tobacco ringspot virus. The presence of the virus was verified by sampling leaves from the experiment and sending them to Agdia Inc. (Elkhart, IN) for verification of the virus

through an ELISA test. At the R2 stage of development (Fehr et al., 1971), soybean plots were visually scored according to the number of plants that exhibited terminal bud death on a scale of 0 (no plants exhibiting bud blight symptoms) to 6 (6 or more plants exhibiting terminal bud death). Data were analyzed by analysis of variance using the Agrobase software (Agronomix Software Inc., Canada).

Bud blight data from the F_4-derived lines were compared with the RFLP data. For the 155 marker loci used, the RFLP class means were compared for the determination of significant difference ($P \leq 0.05$) using an *F*-test from the type III mean squares obtained from the General Linear Model procedure of SAS (SAS Institute, 1988). QTL Cartographer (Windows version 1.21) was also used to more accurately identify the QTL position (Basten et al., 1994; Wang et al., 2001). Zmapqtl Model 6 and a forward regression method were used for composite interval mapping. In the Zmapqtl Model 6, two-control markers and a 5-cM window size were selected.

Confirmation population

A second independent population of 180 F_2-derived lines from the cross of Young x PI416937 was developed (confirmation population). The parents and 180 lines were grown in 1995 at the Univ. of Georgia Plant Sciences Farm, in a randomized complete block design with three replications. Ten entries of each parent were randomized within each replication for a total of 200 entries. The experiment was planted on 5 June 1995 in hill plots with 12 seeds per plot. Hills were planted every 45 cm along rows spaced 76 cm apart. Three weeks after planting, each plot was thinned to six plants per plot. Young

trifoliate leaves from 12 plants of each line were sampled for DNA extraction at the V5 stage of development (Fehr et al., 1971).

Plants were naturally infected with tobacco ringspot virus. The presence of the virus was verified by sampling leaves from the experiment and sending them to Agdia Inc. (Elkhart, IN) for verification of the virus though an ELISA test. At maturity, each hill-plot was visually scored for the percent of plants in each hill with delayed leaf drop and severe discoloration of stems and branches. Data were analyzed by analysis of variance using Agrobase software (Agronomix Software Inc., Canada).

DNA was extracted from leaf tissue using the modified CTAB procedure (Keim et al., 1988). It was then quantified with the UV/VIS Spectrometer (Pekin Elmer) and diluted to 20 ng/µl. DNA bulks consisting of 10 F_2-derived lines without bud blight symptoms and 10 F_2-derived lines with bud blight symptoms were created (Michelmore et al., 1991). A total of four bulks were created, two susceptible (S1 and S2) where each line averaged 100% of plants with bud blight symptoms, and two resistant bulks (R1 and R2) where lines averaged from 0 to 5.7% of plants with symptoms.

Seventeen simple sequence repeat (SSR) markers developed by Cregan et al. (1999) were tested in a 66-cM region of LG F (proposed location of the QTL based on the RFLP data). Fluorescence dye-labeled primers were synthesized (PE-ABI, Foster City, CA) using phosphoramidite chemistry. PCR reactions were prepared using the protocol by Diwan and Cregan (1997). The reactions were performed in a dual 384-well and 96-well GeneAmp PCR System 9700 or a 384-well ABI 877 robotic thermal cycler (PE-ABI, Foster City, CA). The cycling consisted of 1 min at 95°C, followed by 32 cycles of 25 s for denaturation at 94°C, 25 s of annealing at 46°C, and 25 s of extension

at 68°C. At the end of the cycling procedure, the reaction mixtures were held at 4°C. Electrophoresis was run on an ABI-Prism 377 DNA Sequencer (PE-ABI, Foster City, CA) with 12-cm plates at 750 V for 2 h. Lanes were loaded on a 4.8% acrylamide:bisacrylamide (19:1) gel with KLOEHN micro-syringes (Kloehn Ltd., Las Vegas, NV). Genescan (Version 3.0) was used to analyze DNA fragments which were scored with Genotyper (Version 2.1).

Five of the 17 SSR markers were found to be polymorphic for Young and PI416937. The parental DNA and the four bulks were evaluated for their DNA fragments at each polymorphic marker. The population of 180 F_2-derived lines was then evaluated with the five polymorphic markers. The linkage map for the LG F was constructed using the Kosambi map function of MAPMAKER/EXP (Lander et al., 1987). A minimum LOD (Likelihood of odds) of 3.0 and a maximum distance of 37.2 cM was used for establishing linkages among markers. Interval mapping was conducted with the computer program MAPMAKER/QTL (Lincoln et al., 1992) to determine associations between SSR markers on LG F and bud blight incidence. A minimum LOD score of 2.0 was used to declare the presence of a QTL.

Results and discussion

A genetic map of Young x PI416937 has been constructed with 155 polymorphic molecular markers (Lee et al., 1996b). Each RFLP locus on this map was compared with its image in SoyBase (2002) and linkage groups were assigned to their corresponding linkage group on the USDA/Iowa State Univ. public genetic map (Cregan et al., 1999). In

the mapping population that consisted of F_4-derived lines, the mean bud blight rating for Young was 0.0 (based on a 0-6 rating scale), whereas that for PI416937 was 5.6. The F_4-derived progeny lines exhibited a range of bud blight ratings from 0 to 6 (Fig. 1).

The RFLP marker classes (Young/Young vs. PI416937/PI416937) were tested for association of specific marker bands with differences in bud blight resistance among lines using one-way analysis of variance (ANOVA). RFLP markers associated with resistance to bud blight were detected on LG F, LG G, and LG D2 (Table 1). Three of the seven markers represent putative independent QTL. A major locus was identified by marker K644_1 on LG F and explained 82% of the phenotypic variation for bud blight resistance (Table 1). For this marker, the Young allele was associated with bud blight resistance. In addition, two minor QTL were detected on LG G and D2 that explained 9% and 12% of the variation, respectively. For these loci, the PI416937 allele was associated with bud blight resistance. The heritability (selection unit = one location, two replications) of bud blight resistance was 93%, suggesting that the major QTL explains most of the genotypic variation. In this population, the RFLP K644_1 is 2.4 cM from the Blt025_1 locus that explained 77% of the variation and 4.2 cM from MP157_1 locus that explained 65% of the variation (Narvel et al., 2002).

To saturate LG F that harbors the major QTL, four SSR markers (Satt252, Satt516, Satt114, and Satt335) were added. RFLP and SSR molecular data were combined with the phenotypic data for bud blight reaction and composite interval mapping was utilized. Composite interval mapping combines the interval mapping with multiple regression (Zeng, 1994). Multiple regression analysis is useful for determining the most important markers among and within linkage groups and resolving whether

multiple peaks are caused by single or multiple QTL on the same linkage group. The composite interval mapping run with QTL Cartographer Software identified one major QTL on LG F (Fig. 2). No additional QTL on other linkage groups were detected, suggesting that the minor QTL on LG G and D2 identified from one-way ANOVA were most likely Type I errors.

As shown in Figure 2, the major QTL for bud blight resistance is located within a 5.8 cM region, between the RFLP loci Blt025_1 and A069_b, with an LOD score of 64. The high LOD score and the sharpness of the peak position indicate that bud blight resistance is governed by one major gene located on LG F. These results are in agreement with those of *Arabidopsis thaliana*, where genetic mapping studies revealed that tobacco ringspot virus resistance is controlled by a single gene (Lee et al., 1996a).

Yu et al. (1994) used RFLP and SSR markers to identify and map the chromosomal location of the soybean mosaic virus (SMV) resistance gene. They found that two RFLP markers, A186 and K644_1 were closely linked to the SMV resistance gene with distances of 1.5 and 2.1 cM, respectively. This indicates that the SMV resistance gene is in the same chromosomal region and closely linked to the gene for bud blight resistance reported in this study. Both genes were found to be closely linked to the RFLP locus K644_1. In addition, the RFLP markers A186 and K644_1 are linked with a *Phytophthora* root rot resistance gene in soybean (Diers et al., 1992). Furthermore, other studies established that linkage relationships exist between the SMV resistance gene and the genes conditioning resistance to peanut mottle virus (PMV) and peanut stripe virus (PStV) (Roane et al., 1983; Choi et al., 1989). It therefore appears that LG F in soybean encompasses a cluster of virus-resistance genes.

The utilization of RFLP markers for marker-assisted selection (MAS) in a breeding program is difficult because of their low polymorphic content and their high technical demand. Although RFLP-based maps provide a framework for MAS, mapping QTL with SSR markers would greatly facilitate MAS, because SSRs are abundant, highly polymorphic with minimal evidence of clustering, and practical for high throughput application (Akkaya et al., 1995; Maughan et al., 1995; Powell et al., 1996; Diwan and Cregan, 1997). Towards this end, we used SSR markers in a second independent population of Young x PI416937 in order to confirm the position of the QTL for bud blight resistance.

In this study, the confirmation population consisted of F_2-derived lines that possessed from 0 to 100% of their plants exhibiting symptoms of bud blight (Fig. 3). Of the 180 F_2-derived lines, 36 lines had 100% incidence of bud blight (6 plants/hill x 3 reps = 18 plants) and 13 lines had a 0% incidence. PI416937, the susceptible parent, averaged 98% bud blight incidence whereas the resistant parent, Young, averaged 6% incidence (Fig. 3). The range of bud blight incidence among the 10 entries of Young was 0 to 19%, and that of PI416937 was 93 to 100%.

Based on the integrated soybean genetic linkage map (Cregan et al., 1999), we selected 17 SSR markers that had a high probability to map near the RFLP marker K644_1 that is linked to the QTL for bud blight resistance. Five of the 17 SSR markers were found to be polymorphic for Young and PI416937. From bulk segregant analysis, the singular DNA fragment produced from the resistant bulks (R1 and R2) from each of the five polymorphic SSR markers on LG F (Satt114, Satt510, Sat120, Satt335, and Satt362) was the same as the fragment size produced by the resistant parent Young.

Furthermore, the two susceptible bulks produced a singular band, which was the same size as that produced by the susceptible parent PI416937 for all five SSR markers. In addition to the DNA bulks, the entire population of 180 F_2-derived lines was also evaluated with the five polymorphic SSR markers.

Interval mapping confirmed that the QTL conditioning bud blight on LG F was located in the 13.9-cM interval between Satt114 and Satt510 (Fig. 4). According to Mapmaker/QTL, the most likely QTL location was 4 cM from Satt510. On the USDA/Iowa State Univ. soybean linkage map, Satt510 is located 2.3 cM from RFLP marker K644_1, while on the Univ. of Utah map it is within 4.3 cM (Cregan et al., 1999). In the F_4-derived (mapping) population of Young x PI416937, the QTL for bud blight resistance was found to be located between RFLP markers K644_1 and A069_b (Fig. 2). In the F_2-derived population of Young x PI416937, the QTL for bud blight resistance was confirmed and found to be located 4 cM from Satt510 (Fig. 4).

The percent of plants with bud blight symptoms averaged 15.3% for the lines homozygous for the Young fragment at Satt510, 95.1% for the lines homozygous for the PI416937 band, and 62.6% for the lines with both fragments. These data indicate that the bud blight QTL acts in an additive manner. The accuracy of using Satt510 for marker-assisted selection in this population for determining bud blight incidence was evaluated. Figure 5 shows that 47 of 64 lines homozygous for the Young band averaged 20% or less bud blight incidence, whereas 40 of 44 lines homozygous for the PI416937 band averaged 81% or greater bud blight incidence. This suggests that SSR marker Satt510 can be successfully used in marker-assisted selection of bud blight resistance to eliminate lines that are susceptible to bud blight.

The range of bud blight incidence for the lines homozygous for the Young band at Satt510 and homozygous for the PI416937 band did overlap. F_2-derived lines homozygous for the PI416937 band possessed at least 50% of their plants with bud blight. One line homozygous for the Young band contained 75% of its plants with bud blight. This suggests that any lines that are homozygous for the resistant parent band at Satt510 would require additional selection by phenotypic evaluation with the tobacco ringspot virus to identify homogeneous bud blight resistant lines.

The results of this study indicate a major QTL linked to Satt510 on LG F was responsible for bud blight development in this population. The allele from Young at this QTL conditioned resistance. Selection for lines homozygous for the Young allele at Satt510 was successful in eliminating lines highly susceptible to bud blight.

Acknowledgements

This research was supported by funds allocated to Georgia Agricultural Experiment Stations and by grants from the Georgia Agricultural Commodity Commission for Soybeans, Georgia Seed Development Commission and the United Soybean Board.

References

Akkaya, M.S., R.C. Shoemaker, J.E. Specht, A.A. Bhagwat, and P.B. Cregan. 1995. Integration of simple sequence repeat DNA markers into a soybean linkage map. Crop Sci. 35:1439-1445.

Basten, C.J., B.S. Weir, and Z.B. Zeng. 1994. Zmap-a QTL cartographer. p.65-66. *In*: Smith C., J.S. Gavora, B. Benkel, J. Chesnais, W. Fairfull, J.P. Gibson, B.W. Kennedy, and E.B. Burnside (ed.) Proc. 5th World Congress on Genetics Applied to Livestock Production: Computing Strategies and Software. Volume 22. Guelph, Ontario, Canada.

Bergeson, G.B., K.L. Athow, F.A. Laviolette, and Sister Mary Thomasine. 1964. Transmission, movement, and vector relationships of tobacco ringspot virus in soybean. Phytopathology 54:723-728.

Boerma, H.R. and M.A. Mian. 1999. Soybean quantitative trait loci and marker-assisted breeding. p. 68-83. Proc. World Soybean Res. Conf. VI, Chicago, IL.

Burton, J.W., C.A. Brim, and M.F. Young. 1987. Registration of 'Young' soybean. Crop Sci. 27:1093.

Choi, S.H., S.K. Green, and D.R. Lee. 1989. Linkage relationship between two genes conferring resistance to peanut stripe virus and soybean mosaic. Euphytica 44:163-169.

Cregan, P.B., T. Jarvik, A.L. Bush, R.C. Shoemaker, K.G. Lark, A.L. Kahler, T.T. Van Toai, D.G. Lohnes, J. Chung, and J.E. Specht. 1999. An integrated genetic linkage map of the soybean genome. Crop Sci. 39:1464-1490.

Crittenden, H.W., K.M. Hastings, and D.M. Moore. 1966. Soybean losses caused by tobacco ringspot virus. Plant Dis. Reptr. 50:910-913.

Demski, J.W. and C.W. Kuhn. 1999. Tobacco ringspot. p. 66-68. *In* G.L. Hartman, J.B. Sinclair, and J.C. Rupe (ed.) Compendium of soybean diseases. Fourth edition. American Phytopathological Society, St. Paul, MN.

Diers, B.W., L. Mansur, J. Imsande, and R.C. Shoemaker. 1992. Mapping *Phytophthora* resistance loci in soybean with restriction fragment length polymorphism markers. Crop Sci. 32:377-383.

Diwan, N. and P.B. Cregan. 1997. Automated sizing of fluorescent-labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. Theor. Appl. Genet. 95:723-733.

Dunleavy, J.M. 1957. The grasshopper as a vector of tobacco ringspot virus in soybean. Phytopathology 47:681-682.

Fehr, W.R., C.E. Caviness, D.T. Burmood, and J.S. Pennington. 1971. Stage of development descriptions for soybeans, *Glycine max* (L.) Merrill. Crop Sci. 11:929-931.

Flor, H.H. 1947. Inheritance of reaction to rust in flax. J. Agric. Res. 74:241-262

Holloway, J.L., and S.J. Knapp. 1993. Gmendel 3.0 user's guide. Dept. of Crop and Soil Sciences. Oregon State University.

Kahn, R.P., and F.M. Latterell. 1955. Symptoms of bud-blight of soybeans caused by the tobacco and tomato ringspot viruses. Phytopathology 45:500-502.

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

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

Lee, J-M., G.L. Hartman, L.L. Domier, and A.F. Bent. 1996a. Identification and map location of *TTR1*, a single locus in *Arabidopsis thaliana* that confers tolerance to tobacco ringspot nepovirus. Molecular Plant-Microbe Interactions 9:729-735.

Lee, S.H., M.A. Bailey, M.A.R. Mian, T.E. Carter, Jr., D.A. Ashley, R.S. Hussey, W.A. Parrott, and H.R. Boerma. 1996b. Molecular markers associated with soybean plant height, lodging, and maturity across locations. Crop Sci. 36:728-735.

Lincoln, S., M. Daly, and E. Lander. 1992. Mapping genes controlling quantitative traits with MAPMAKER/QTL. Whitehead Institute Tech. Rept. 2nd ed. Whitehead Institute, Cambridge, MA.

Maughan, P.J., M.A. Saghai-Maroof, and G.R. Buss. 1995. Microsatellite and amplified sequence length polymorphisms in cultivated and wild soybean. Genome 38:715-723.

Michelmore, R.W., I. Paran, and R.V. Kesseli. 1991. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc. Natl. Acad. Sci. (USA) 88:9828-9832.

Mitchell, S.E, S. Kresovich, C.A. Jester, C.J. Hernandez, and A.K. Szewc-McFadden. 1997. Application of multiplex PCR and fluorescence-based, semi-automated

allele sizing technology for genotyping plant genetic resources. Crop Sci. 37:617-624.

Narvel, J.M., T.E. Carter, Jr., L.R. Jakkula, J. Alvernaz, M.A. Bailey, M.A.R. Mian, S.H. Lee, G.J. Lee, and H.R. Boerma. 2002. Registration of NC113 soybean mapping population. Crop Sci. 42:(in press).

Orellana, R.G. 1981. Resistance to bud blight in introductions from the germ plasm of wild soybean. Plant Dis. 65:594-595.

Powell, W., M. Morgante, C. Andre, M. Hanafey, J. Vogel, S. Tingey, and A. Rafalski. 1996.The comparison of RFLP, RAPD, AFLP, and SSR (microsatellite) markers for germplasm analysis. Mol. Breeding 2:225-238.

Roane, C.W., S.A. Tolin, and G.R. Buss. 1983. Inheritance of reaction to two viruses in the cross York x Lee 68. J. Hered. 74:289-291.

Rongwen, J., M.S. Akkaya, A.A. Bhagwat, U. Lavi, and P.B. Cregan. 1995. The use of microsatellite DNA markers for soybean genotype identification. Theor. Appl. Genet. 90:43-48.

SAS. 1988. SAS/STAT user's guide, version 6.03. SAS Institute, Cary, NC.

SoyBase. 2002. (http://genome.cornell.edu/cgi-bin/WebAce/webace?db=soybase).

Tu, J.C. 1986. Strains of tobacco ringspot virus isolated from soybean (*Glycine max*) in southwestern Ontario (Canada). Can. J. Plant Sci. 66:491-498.

Wang, S., C.J. Basten, and Z.B. Zeng. 2001. Windows QTL cartographer, Version 1.20. Department of Statistics, North Carolina State University, Raleigh, NC.

Yu, Y.G., M.A. Saghai Maroof, G.R. Buss, P.J. Maughan, and S.A. Tolin. 1994. RFLP and microsatellite mapping of a gene for soybean mosaic virus resistance. Phytopathology 84:60-64.

Zeng, Z. 1994. Precision mapping of quantitative trait loci. Genetics 136:1457-1468.

Table 3.1. RFLP markers identified by analysis of variance and associated with resistance to bud blight in the F_4-derived soybean population of Young x PI416937.

RFLP locus	Linkage Group	Allelic means †		*P*	R^2
		Y/Y	PI/PI		
		----- rating ‡ -----			%
MP157_1	F	0.8	5.0	0.0001	65
A186_1	F	0.8	5.1	0.0001	68
K644_1	F	0.5	5.2	0.0001	82
Blt025_1	F	0.6	5.2	0.0001	77
A069_b	F	0.5	5.0	0.0001	74
L183_1	G	3.6	2.0	0.0025	9
K258_2	D2	3.6	1.8	0.0005	12

† Y/Y = homozygous for the band from Young; PI/PI = homozygous for the band from PI416937.

‡ Rating: 0 = no plants exhibiting bud blight symptoms to 6 = 6 or more plants exhibiting terminal bud death.

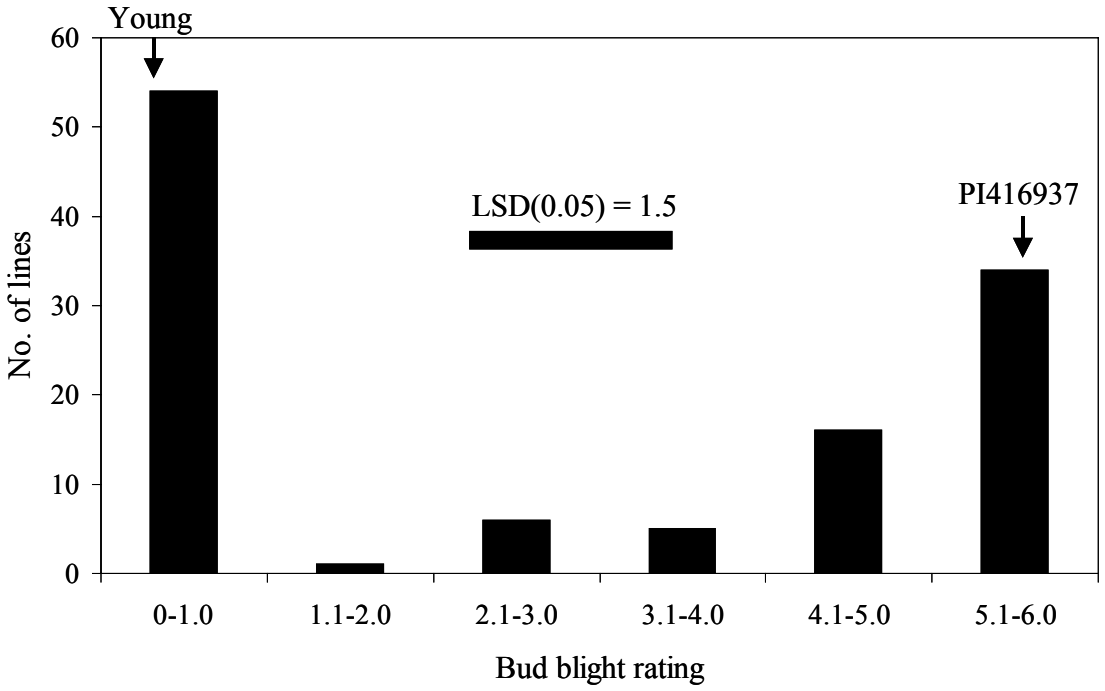

Figure 3.1. Distribution of F_4-derived lines of the mapping population of Young x PI416937 for bud blight rating (Rating: 0 = no plants exhibiting bud blight symptoms to 6 = 6 or more plants exhibiting terminal bud death).

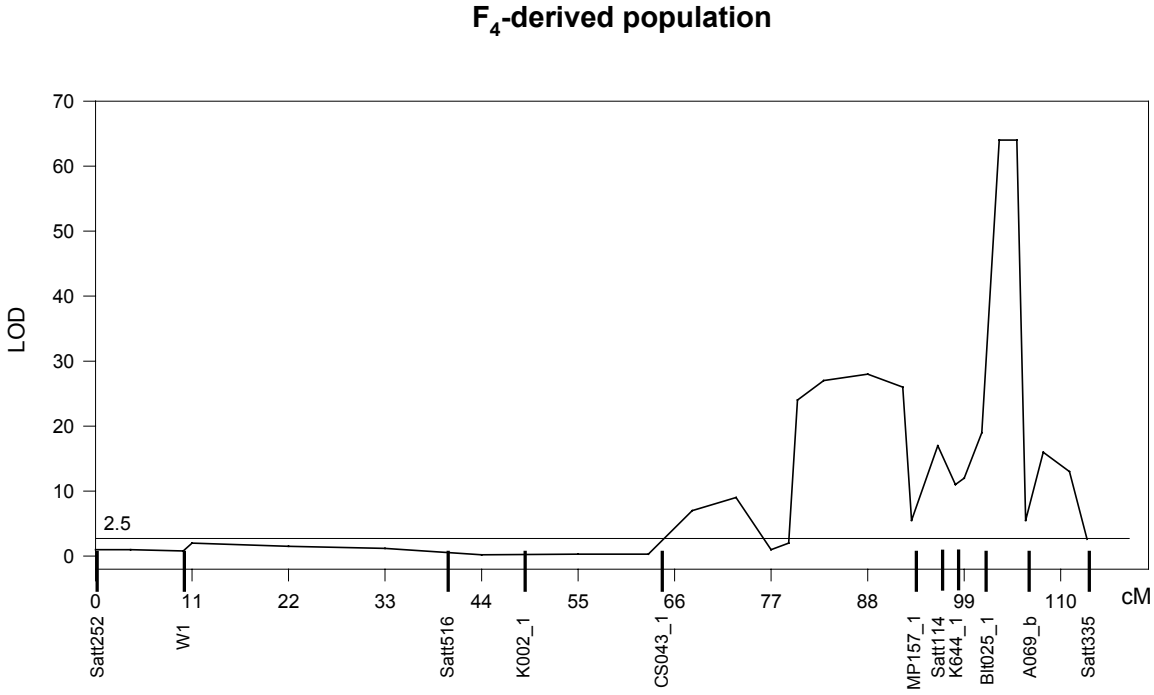

Figure 3.2. QTL-likelihood plot for bud blight on Linkage Group F for the Young x PI416937 F_4-derived soybean population based on composite interval mapping run with QTL Cartographer. The significance threshold is indicated by a line at LOD = 2.5.

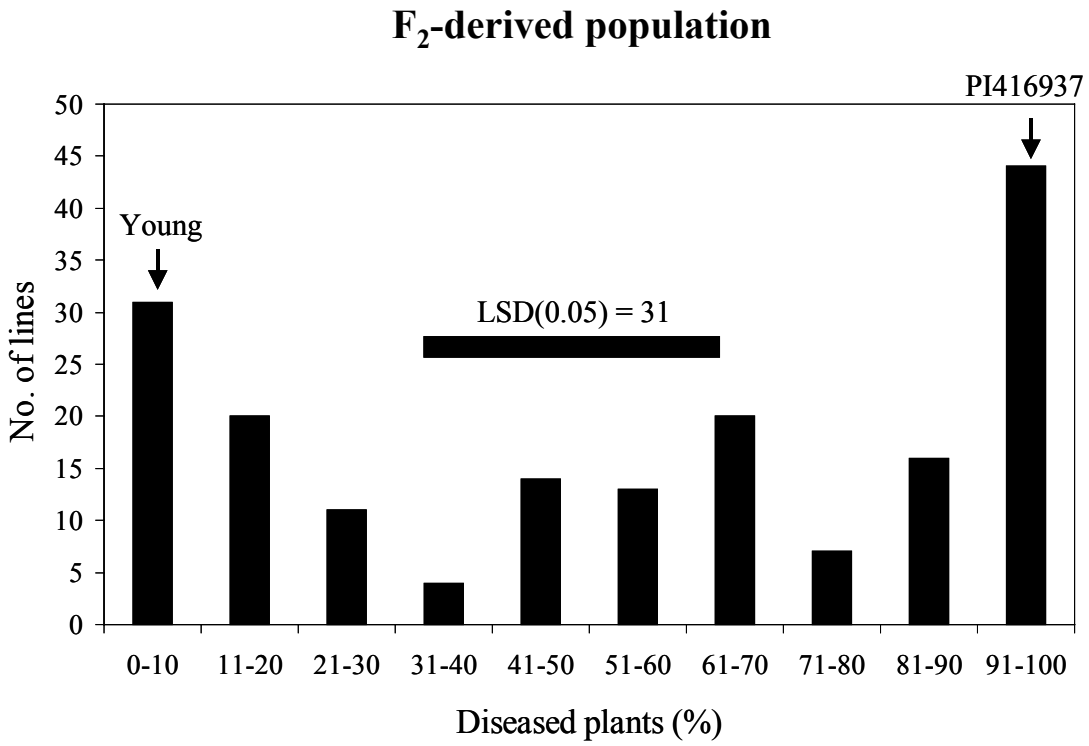

Figure 3.3. Distribution of F_2-derived lines of the confirmation population of Young x PI416937 for percent of plants exhibiting bud blight.

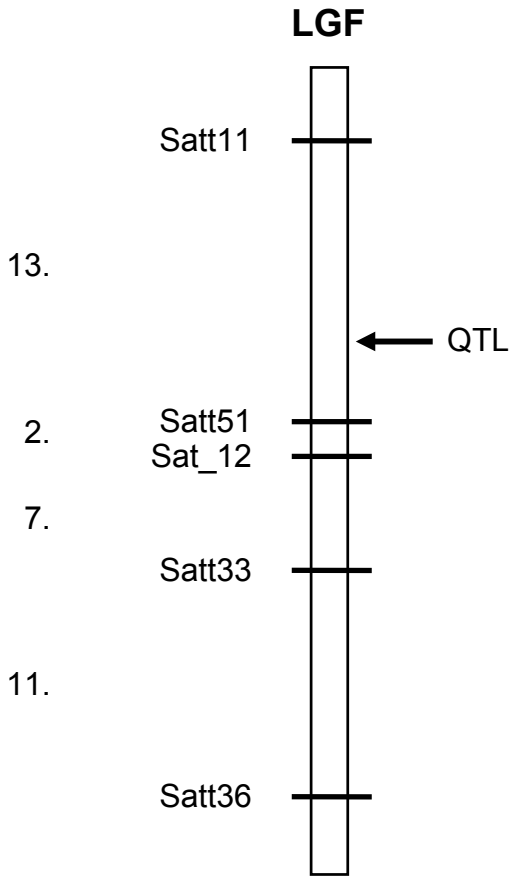

Figure 3.4. Linkage Group F for the Young x PI416937 F_2-derived soybean population showing the five polymorphic SSR markers and their distance in cM. Location of RFLP marker K644_1 is approximated based on Cregan et al. (1999) and Narvel et al. (2002).

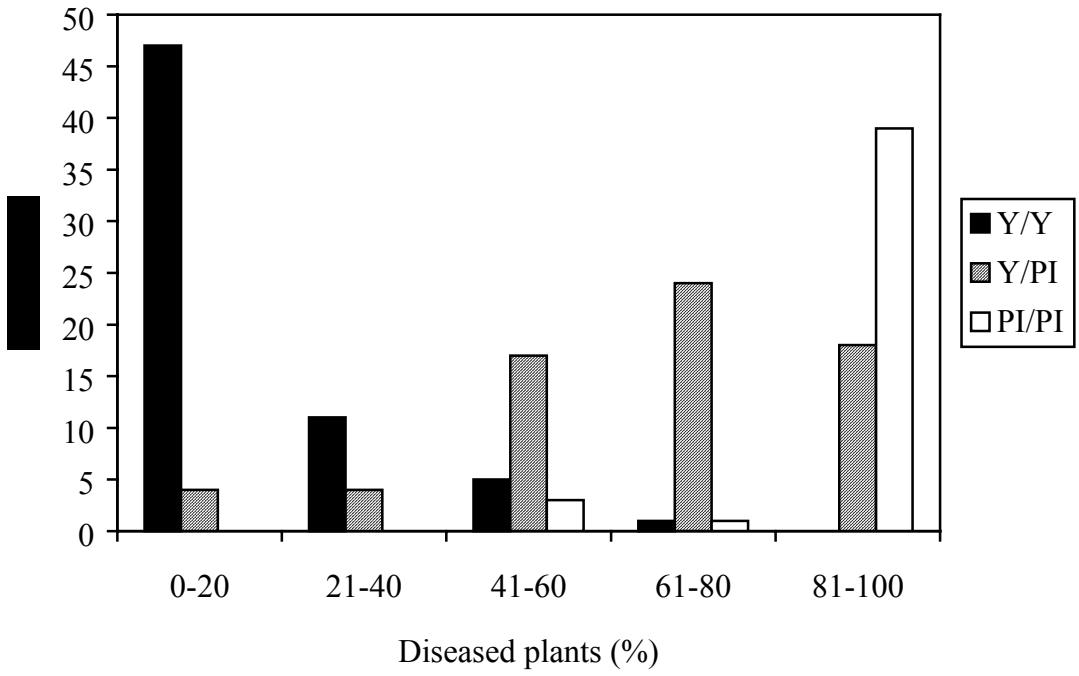

Figure 3.5. Mean incidence of bud blight in 180 F_2-derived soybean lines homozygous for the Young (Y/Y) band, homozygous for the PI416937 (PI/PI) band, and heterogeneous (Y/PI) for bands at Satt510.

CHAPTER 4

CONSISTENCY OF QUANTITATIVE TRAIT LOCI FOR SEED COMPOSITION AND SEED WEIGHT IN A SOYBEAN POPULATION[1]

---

[1]V.A. Fasoula and H.R. Boerma. To be submitted to Theoretical and Applied Genetics.

Abstract

The technological advancements in molecular biology have created the opportunity for genetic dissection of complex traits into a set of discrete loci. RFLP markers have been used extensively to map the genomic location of soybean quantitative trait loci (QTL) for many agronomic, physiological, and seed composition traits. In soybean and other crops, there is limited and imprecise information confirming the previously reported QTLs. The objectives of this study were to: (1) utilize an independent F_2-derived soybean, *Glycine max* (L.) Merr., population of PI97100 x Coker 237 to confirm previously reported RFLP markers associated with seed protein, seed oil, and seed weight, mapped in a F_2-derived population created from the same parents, and (2) investigate the consistency of QTLs across three different environments. Single-factor analysis of variance (ANOVA) was used to confirm the RFLP loci that are significantly ($P \leq 0.05$) associated with seed composition and seed weight. Two (E/A454-1 and UNK/A132-4) out of four previously described QTLs for seed protein, two (C1/A063-1 and H/A566-2) out of three QTLs for seed oil content, and one (D2/A257-1) out of three QTLs for seed weight were confirmed in the independent population. Thus, 50% of the QTLs detected in the original mapping studies were verified in the new population of PI97100 x Coker 237. In general, the confirmed QTLs were detected based on the mean phenotypic data across the three environments as well as within each environment. Therefore, these QTLs were consistent across environments and have been confirmed across two independent populations. Some of the previously identified QTLs may have been environmentally sensitive in that they were only identified in one of two

environments in the original mapping study. Other QTLs that were not detected in this study may have been erroneously declared significant in the original population (Type I error) or they may have been specific for the sample of lines used in the original population. These results confirm the necessity of mapping quantitative trait loci in multiple environments and parallel populations before utilizing them in a plant improvement program.

Introduction

The technological advancements in molecular biology have created the opportunity for genetic dissection of complex traits into a set of discrete loci (Tanksley et al. 1989). The ability to utilize DNA markers to identify the genomic location of plant genes has played an important role in revolutionizing the science of plant breeding and genetics. In soybean, RFLP (restriction fragment length polymorphism) markers were introduced in the late 1980s (Apuya et al. 1988; Keim et al. 1989). RFLP markers have been used extensively to map the genomic location of quantitative trait loci (QTL) for many agronomic, physiological, and seed composition traits in soybean (Boerma 2000).

Although more than 300 QTLs have been reported for the various quantitative traits, limited research exists on confirmation of the reported soybean QTLs and the results from these studies have been mixed. For example, Diers et al. (1992a) reported that none of the identified QTLs conditioning iron deficiency chlorosis in soybean were effective for divergent selection among lines from the same cross that were not used in the original QTL mapping study. Mudge et al. (1997) found that a single SSR (simple sequence repeat) marker was 95% accurate in predicting resistance to soybean cyst nematode. Moreover, when two SSR markers that flanked the QTL were used, the accuracy of predicting the resistant phenotype was increased to 98%. In another study, Li et al. (2001) reported that marker assisted selection at two QTL conditioning resistance to root-knot nematode was successful in identifying resistant lines in an independently derived soybean population.

Brummer et al. (1997) identified QTLs for soybean seed protein and oil content using eight distinct populations. They reported that the phenotypic effect of some QTLs was sensitive to the environment in which they were evaluated and they detected both environmentally sensitive and stable QTLs. Harris et al. (2001) utilized SSR markers in soybean in order to verify previously reported QTLs associated with seed protein, seed oil, and seed weight in a newly created soybean population derived from the same parents. In their study, SSR markers were selected from the consensus soybean map in the same genomic region as the RFLP markers identified in the original mapping study. They were able to confirm only 33% of the QTLs reported in the original mapping study.

In maize (*Zea mays* L.), results from three independent experiments repeated in the same genetic background revealed that the QTLs identified were not consistent (Beavis et al. 1994; Beavis 1994). Confounding factors such as population structure, sources of parental lines, different sets of environments, and sampling of progeny have been reported as possible causes for this discrepancy (Beavis 1994). In another maize study, Ajmone-Marsan et al. (1996) evaluated previously identified QTLs for grain yield in an independent sample drawn from the same population. They found that two QTLs were consistent with those detected in the previous experiments, but two QTLs identified in the first sample remained undetected in the independent sample.

Most qualitative genetic studies in soybean require a hypothesis generation and usually a second or confirming generation in order to assign a gene symbol. However, this confirmation step has not been required in QTL mapping studies in soybean or other species (Boerma and Mian 1999). Since there is limited and conflicting information confirming the reported QTLs in soybean, it is important to conduct confirmation

experiments prior to the development of breeding strategies based on unconfirmed QTL reported in the literature. Furthermore, a number of QTL mapping studies have not used multiple environments or populations for the collection of phenotypic data. Studies conducted in a single environment are likely to underestimate the number of QTLs which can influence a trait (Paterson et al. 1991).

In soybean, a number of important QTL mapping studies have been conducted for seed composition traits. Soybean seed is a major source of protein for animal feed and oil for human consumption. Simultaneous increases in protein and oil content can proceed only to a limited extent since most experimental data show that protein and oil content are negatively correlated (Burton 1987). Intense breeding efforts have resulted in the selection of two types of seed composition, those with a higher percentage of protein content and those with a higher percentage of oil content (Miller and Fehr 1979; Brim and Burton 1979; Burton and Brim 1981; Burton 1985; Wilcox 1985).

Soybean mapping studies have identified a number of QTLs for seed protein and oil content (Diers et al. 1992b; Lee et al. 1996b; Mansur et al. 1993; Mansur et al. 1996; Brummer et al. 1997; Qiu et al. 1999; Sebolt et al. 2000). Lee et al. (1996b) reported RFLP markers associated with seed protein and oil in 111 F_2-derived lines of the PI97100 x Coker 237 soybean population. Six RFLP markers were associated with seed protein, four of which were independent QTLs residing on different linkage groups. The four independent markers explained 46% of the total phenotypic variation for seed protein. The heritability (assuming a selection unit of two locations, three replications/location) for seed protein in this population was 83%. In addition, Lee et al. (1996) detected three independent RFLP markers associated with seed oil content. The amount of phenotypic

variation explained by these markers was 40%, whereas the heritability for seed oil in this population was 75%.

Seed weight, measured as mass per seed, is an important yield component in soybean and is generally positively correlated with seed yield (Burton 1987). Soybean cultivars with either very small or very large seed sizes are used in the production of many specialty human foods. The demand for these food-type soybeans is steadily increasing in the global market at a rate of 3 to 5% per year. Sales of the food-type soybeans have increased by 450% in the last 18 years (Wilson 1999). Mapping studies in soybean have identified QTLs for seed weight in different populations (Mansur et al. 1993; Mansur et al. 1996; Mian et al. 1996; Maughan et al. 1996). Mian et al. (1996) reported RFLP markers associated with seed weight in the PI97100 x Coker 237 population. Eight RFLP markers were found to be associated with seed weight and explained 74% of the total phenotypic variation. The heritability of seed weight in this population was 91%. Thus, the majority of the phenotypic variation could be explained by the eight RFLP markers (Mian et al. 1996).

The objective of this study was to utilize an independent F_2-derived population of PI97100 x Coker 237 to confirm previously reported RFLP markers associated with QTLs for seed protein, seed oil, and seed weight. A second objective was to investigate the consistency of the QTLs across different environments.

Materials and Methods

A soybean population of 180 F_2-derived lines from the cross of PI97100 x Coker 237 was developed. PI97100 possesses an indeterminate growth habit while Coker 237 has a determinate growth habit. The cross was created in the summer of 1993. The F_1 generation was grown in the greenhouse in Athens, GA, and the F_2 generation was grown at the Univ. of Georgia Plant Sciences Farm in Athens, GA in 1994. At maturity, 180 randomly selected plants were individually harvested to create F_2-derived lines. The parents and the 180 F_2-derived lines were planted on 5 June 1995 at the Univ. of Georgia Plant Sciences Farm (Athens 95), in a randomized complete block design with three replications. Ten entries of each parent were randomized within each replication to create a total of 200 entries. The experiment was planted in hill plots with 12 seeds per plot. Hill plots were planted every 45 cm, along rows spaced 76 cm apart. Three weeks after planting each hill was thinned to six plants per plot. At maturity, the plants in each hill plot were cut by hand and threshed with a plot combine.

In 1996, the F_2-derived lines of the PI97100 x Coker 237 population were grown at the Univ. of Georgia Plant Sciences Farm in Athens, GA (Athens 96) and the Univ. of Georgia Southwest Branch Experiment Station in Plains, GA (Plains 96). The experimental unit for each entry was two 4m rows spaced 76 cm apart at Athens and two 7m rows spaced 76 cm apart at Plains. These lines were randomly assigned to four tests of 44 lines for a total of 176 F_2-derived lines (four of the original 180 F_2-derived lines were not included). In addition to the 44 F_2-derived lines, each test included three entries of PI97100 and Coker 237 (total of 50 entries per test). Each test was planted in a

randomized complete block design with two replications. At maturity each plot was harvested.

For protein and oil content determination, a 50-g seed sample from each plot was sent to the USDA-ARS National Center for Agricultural Utilization Research at Peoria, IL for seed composition analysis. An 18 to 20-g sample of seed was analyzed for protein and oil composition with a model 1255 Infratec NIR food and feed grain analyzer. The protein and oil values were converted to moisture-free basis. The seed weight for each plot was determined based on a 100-seed sample. The mean phenotypic data for protein, oil, and seed weight from the four tests did not significantly differ based on t-tests using the error variance from each test. Thus, the data were not standardized across tests.

Data were analyzed by analysis of variance using the Agrobase software (Agronomix Software Inc., Canada). For the statistical model over locations in 1996 and over the three environments (Athens 95, Athens 96, and Plains 96), locations, environments, replications, and genotypes were considered random effects.

Young trifoliate leaves from 12 plants of each line (2 hills) were sampled for DNA extraction from the 1995 hill plot experiment. DNA isolation, restriction enzyme digestion, electrophoresis, southern blotting, and hybridization procedures were performed according to Lee et al. (1996a,b). RFLP markers associated with independent protein, oil, and seed weight QTLs reported by Mian et al. (1996) and Lee et al. (1996b) were used to confirm these QTLs in the newly-created population of PI97100 x Coker 237. The following RFLP loci were evaluated for seed protein content: E/A454-1, K/A065-1, UNK/A132-4, H/A566-2; seed oil: C1/A063-1, G/L154-2, H/A566-2; and

seed weight: D2/A257-1, G/A235-1, M/Cr529-1 (where E/A454-1 refers to RFLP marker A454-1 on Linkage Group E).

The phenotypic data from the F_2-derived lines were analyzed for the appropriate RFLP markers. Single-factor analysis of variance (ANOVA) was used to determine the significance ($P \leq 0.05$) among the RFLP genotypic class means using an *F*-test from the Type III mean squares obtained from the GLM procedure (SAS Institute, 1989). The mean seed protein content, oil content, and seed weight across years and locations as well as across individual environments, were compared for the lines homozygous for the PI97100 RFLP band and the lines homozygous for the Coker 237 RFLP band at each QTL. Previously reported QTLs were assumed to be confirmed if the means of these two groups were significantly different ($P \leq 0.05$) and the parental alleles produced a similar effect as in the original mapping study.

Results

Seed protein

The mean phenotypic data for seed protein, seed oil, and seed weight from the four individual tests at each location in 1996 were similar. Thus, it was unnecessary to adjust the line means in the separate tests. The protein content of the 180 F_2-derived lines showed continuous variation (Fig. 1). Combined analysis over three environments (Athens 95, Athens 96, and Plains 96) indicated that PI97100 and Coker 237 differed by 42 g/kg in seed protein content, with PI97100 having 10% higher protein content than

Coker 237. The seed protein of the progeny ranged from 423 g/kg to 478 g/kg and the mean protein content of the population was 450 g/kg (Fig. 1).

Four independent QTLs for seed protein were detected in the original F_2-derived population of PI97100 x Coker 237 that consisted of 111 lines (Table 1) (Lee et al. 1996b). We utilized these RFLP markers to verify the protein QTLs in our F_2-derived PI97100 x Coker 237 population that consisted of 180 lines. Single-factor analysis of variance (ANOVA) revealed that two RFLP loci were confirmed on the basis of combined analysis over the three environments (Table 2). RFLP markers E/A454-1 and UNK/A132-4 were found to be associated with seed protein ($P \leq 0.0000$ and $P \leq 0.0116$, respectively). The QTL at E/A454-1 was significant in all three environments (Athens 95, Athens 96, and Plains 96) whereas the UNK/A132-4 QTL was significant ($P \leq 0.05$) in Plains 96 and approached significance ($P \leq 0.1$) in Athens 95 and Athens 96 (Table 2).

As in the original mapping study (Table 1; Lee et al. 1996b), we found that for E/A454-1 locus, the PI97100 allele was associated with increased protein, whereas for UNK/A132-4 locus, the Coker 237 allele was associated with increased protein. Based on the combined data, the E/A454-1 QTL accounted for 12.3% of the total phenotypic variation for seed protein and the UNK/A132-4 QTL accounted for 5.6% of the total phenotypic variation. The H/A566-2 locus was not detected to be significant in any environment in our study. In addition, the previously identified K/A065-1 marker was not detected in the independent population of PI97100 x Coker 237. In the original mapping study, the K/A065-1 marker was detected only in one of the two environments evaluated, but it had a large effect (R^2=21%), and was also detected in the combined analysis of the data (Lee et al. 1996b). Most likely, the effect of this QTL was dependent on the specific

environment that it was detected (environmentally sensitive QTL). These data are in agreement with the results of Brummer et al. (1997) who identified QTLs for seed protein and oil using eight distinct soybean populations. In all eight populations, they detected both environmentally sensitive and environmentally stable QTLs for seed protein and oil content.

Seed oil

The seed oil content of the F_2-derived lines showed continuous variation (Fig. 2). Combined analysis over the three environments (Athens 95, Athens 96, and Plains 96) showed that PI97100 and Coker 237 differed by 28 g/kg in seed oil content, with Coker 237 having 16% higher oil content than PI97100. The seed oil of the progeny lines ranged from 169 g/kg to 197 g/kg and the mean oil content of the population was 185 g/kg (Fig. 2).

Three independent QTLs for seed oil were detected in the original mapping population of PI97100 x Coker 237 (Table 1) (Lee et al. 1996b). These previously reported RFLP markers were used to verify the seed oil QTLs in the newly derived PI97100 x Coker 237 population. Single-factor ANOVA revealed that two of the QTLs were detected on the basis of combined analysis over the three environments (Table 2). RFLP markers C1/A063-1 and H/A566-2 were confirmed to be associated with seed oil content ($P \leq 0.0011$ and $P \leq 0.0008$, respectively) in the independent population of PI97100 x Coker 237. Both QTLs were significant ($P \leq 0.05$) in the three individual environments (Athens 95, Athens 96, and Plains 96) as well as in the combined analysis

over all the environments. Consistent with the original mapping study (Table 1) (Lee et al. 1996b), we found that the PI97100 allele was associated with increased oil at the C1/A063-1 locus, whereas for the H/A566-2 locus, the Coker 237 allele was associated with increased oil. The C1/A063-1 and H/A566-2 QTLs each accounted for 8% of the total phenotypic variation for seed oil (Table 2).

The G/L154-2 oil QTL was not detected to be significant in our study. Lee et al. (1996b) reported that at the G/L154-2 locus, lines with both marker bands had a higher seed oil percentage than homozygous lines, suggesting overdominance or pseudo-overdominance (Table 1). Most likely, the G/L154-2 locus exhibited pseudo-overdominance, i.e., nonadditive allelic action of two genes linked in repulsion phase (Fasoula and Fasoula 1997). The G/L154-2 locus explained 21% of the total phenotypic variation, but it was detected in only one environment, therefore it was not an environmentally stable QTL. In our newly derived population of PI97100 x Coker 237, the G/L154-2 locus was not detected to be significant in any environment and there was also no evidence of pseudo-overdominance (Table 2). Most likely, the QTL associated with the G/L154-2 marker was dependent on the specific environment it was initially detected (unstable QTL). Alternatively, the effect of this QTL may have been dependent upon the specific sample of lines used in the original population or may have been erroneously declared significant in the original population (Type I error).

In soybean, seed protein and oil contents have been reported to be negatively correlated (Burton 1987). In this experiment, negative correlations between oil and protein were observed in both 1995 and 1996 ($r = -0.64$ and $r = -0.55$, respectively). The negative association for protein and oil contents is in agreement with earlier studies

(Johnson and Bernard 1962; Kwon and Torrie 1964; Smith and Weber 1968). In some mapping studies, the association between these two traits was explained by QTLs conditioning both traits (Diers et al. 1992b; Mansur et al. 1993). We analyzed all the RFLP markers for protein or oil in Table 2 for both seed protein and seed oil content and we did not detect any common QTL for protein and oil content (data not shown).

The QTL linked to H/A566-2 locus was detected to be associated with both traits in the original mapping population (Lee et al. 1996b). In our study, H/A566-2 was detected to be significant only for seed oil content. The C1/A063-1 marker was detected to be associated with seed oil content in both the original population and the independently derived population (Table 1 and 2). This marker has also been reported to be significantly associated with seed protein content in other populations (Brummer et al. 1997). In addition, E/A454-1 marker was significantly associated with seed protein content in both the original and the independent population of PI97100 x Coker 237 (Tables 1 and 2). This marker is also reported to be associated with seed oil content (Diers et al., 1992b).

Seed weight

The mean seed weight of the 176 lines in the PI97100 x Coker 237 population showed a continuous distribution (Fig. 3). Combined analysis over the three environments (Athens 95, Athens 96, and Plains 96) showed that Coker 237 had 4% larger seed weight than PI97100, but the difference was not statistically significant ($P > 0.05$). The seed weight of the F_2-derived lines ranged from 133 mg/seed to 196 mg/seed

and the mean seed weight of the population was 161 mg/seed. The progeny exhibited transgressive segregation for both larger and smaller seed weight than the parents (Fig. 3).

Eight independent QTLs for seed weight have been detected in the original mapping population of PI97100 x Coker 237 (Mian et al. 1996). The R^2 value for these QTLs ranged from 5% to 11%. We chose to evaluate three QTLs that were detected in all three locations and had an R^2 value $\geq$ 8% (Table 1). Based on single-factor ANOVA, none of the three RFLP loci were significantly associated with seed weight in the independently derived soybean population that consisted of 180 F_2-derived lines (Table 2). However, when a t-test was used to compare the two homozygous classes, the mean of lines possessing the D2/A257-1 allele from Coker 237 was significantly different ($P \leq 0.05$) from the mean of lines possessing the D2/A257-1 allele from PI97100. Consistent with the original mapping study (Table 1), the Coker 237 allele was associated with larger seed weight. Previously reported RFLP markers G/A235-1 and M/Cr529-1 could not be confirmed to be associated with seed weight in any of the three individual environments or in the combined data.

Several explanations are possible for our inability to confirm the seed weight QTLs. The effect of these putative QTLs was dependent upon the environments in which they were originally tested. In other words, these QTLs may be environmentally sensitive. It is also possible that the effect of these QTLs was dependant upon the specific sample of lines in the original population (i.e., limited population size). Alternatively, these markers may have been erroneously declared significant in the original mapping population, when in reality they are not (Type I error). These results confirm the

necessity of mapping quantitative trait loci in multiple environments and populations before utilizing them in plant improvement programs.

Discussion

In the independent soybean population of PI97100 x Coker 237, we were able to confirm two out of four previously describe QTLs for seed protein, two out of three QTLs for seed oil content, and one out of three QTLs for seed weight. Thus, 50% of the QTLs detected in the original mapping studies were confirmed in the new population of PI97100 x Coker 237. It is important to note that the verified QTLs were generally detected across environments as well as within each environment (Table 2). Therefore, we conclude that these QTLs were consistent across environments and repeatable across two independent populations created from the same parents.

Some QTLs (i.e., K/A065-1 and G/L154-2) were detected in only one location in the original studies (Lee et al. 1996; Mian et al. 1996). These QTLs were not detected in our population; therefore, we conclude that they are likely environmentally sensitive QTLs. In addition, two QTLs for seed weight that were detected in all locations in the original study were not verified in our population. These QTLs may have been erroneously declared significant in the original population (Type I error) or they may have been detected due to the limited or specific sampling of lines in the original population (111 F_2-derived lines). Lande and Thompson (1990) reported that the QTL effects estimated from the same data used for QTL mapping were generally overestimated. Computer simulations have demonstrated that the upward bias can be

quite severe (Beavis 1994). Melchinger et al. (1998) reported that estimates of the phenotypic and genetic variance explained by QTLs were considerably reduced when derived from an independent validation sample as opposed to estimates from the calibration sample of the same population used to map the QTLs.

Brummer et al. (1997) identified QTLs for soybean seed protein and oil content using eight distinct populations and reported that some QTLs were sensitive to the environment they were initially detected. Harris et al. (2001) utilized SSR markers in soybean in order to verify previously reported QTLs associated with seed protein, seed oil, and seed weight in a newly created F_2-derived soybean population and confirmed only 33% of the QTLs. In their study, they utilized SSR markers that were selected from the consensus soybean map in the same genomic region as the RFLP markers identified in the original mapping study. Other studies in soybean have provided mixed results regarding the confirmation of reported QTLs (Diers et al. 1992b; Mudge et al. 1997; Li et al. 2001).

In maize, results from three independent experiments repeated in the same genetic background revealed that the identified QTLs were not consistent (Beavis et al. 1994; Beavis 1994). In another maize study, Ajmone-Marsan et al. (1996) evaluated previously identified QTLs for grain yield in an independent sample drawn from the same population. They found that two QTLs were consistent with those detected in the previous experiments, but two QTLs identified in the first sample remained undetected. Melchinger et al. (1998) genotyped two independent samples from the same F_2 maize population using RFLP markers. QTLs for grain yield and other agronomically important traits were mapped in both samples. For all traits, they detected a total of 107 QTLs from

the first sample and 39 QTLs from the second independent sample. They found that only 20 QTLs were common in both population samples.

Boerma and Mian (1999) reported that there is very limited information confirming the reported QTLs in soybean and that this confirmation step has not been required in QTL mapping studies. On the contrary, most qualitative genetic studies require a hypothesis generation and usually a second or confirming generation in order to assign a gene symbol. The precise identification of QTLs is necessary for successful application of marker-assisted selection (MAS) in plant improvement programs. Simply demonstrating that a complex trait can be dissected into QTL and mapped to approximate genomic locations using DNA markers is inadequate (Young 1999). Our data indicate that in addition to improved phenotypic data collection, larger population sizes, and multiple replications and environments, the precise identification of QTLs requires independent verification through advanced generations or parallel populations.

Acknowledgements

This research was supported by funds allocated to Georgia Agricultural Experiment Stations and by grants from the Georgia Agricultural Commodity Commission for Soybeans, the Georgia Seed Development Commission, and the United Soybean Board.

References

Ajmone-Marsan P, Monfredini G, Brandolini A, Melchinger AE, Garay G, Motto M (1996) Identification of QTL for grain yield in an elite hybrid of maize: repeatability of map position and effects in independent samples derived from the same population. Maydica 41:49-50

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

Beavis WD (1994) The power and deceit of QTL experiments: lessons from comparative QTL studies. In: Proc of the Forty-nineth Annual Corn and Sorghum Industry Research Conference ASTA, Washington DC, pp 250-266

Beavis WD, Smith OS, Grant D, Fincher R (1994) Identification of quantitative trait loci using a small sample of topcrossed and F4 progeny of maize. Crop Sci 34:882-896

Boerma HR (2000) DNA markers: are they worth the cost? Proc American Seed Trade Association, Soybean Seed Research Conf. Chicago, Illinois, Dec. 1999

Boerma HR, Mian MAR (1999) Soybean quantitative trait loci and marker-assisted breeding. World Soybean Res Conf VI Proc, Chicago, Illinois, pp 68-83

Brim CA, Burton JW (1979) Recurrent selection in soybean. II. Selection for increased percent protein in seeds. Crop Sci 19:494-498

Brummer EC, Graef GL, Orf J, Wilcox JR, Shoemaker RC (1997) Mapping QTL for seed protein and oil content in eight soybean populations. Crop Sci 37:370-378

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

Burton JW (1987) Quantitative genetics: Results relevant to soybean breeding. In: Wilcox JR (ed) Soybeans: Improvement, production and uses. 2nd ed. Agron. Monogr. 16. ASA, CSSA, and SSSA, Madison, WI.

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

Diers BW, Cianzio SR, Shoemaker RC (1992a) Possible identification of quantitative trait loci affecting iron efficiency in soybean. J Plant Nutrition 15:2127-2136

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

Fasoula DA, Fasoula VA (1997) Gene action and plant breeding. Plant Breed Rev 15:315-374

Harris DK (2001) Genetic improvement of soybean seed trait and resistance to bud blight and root-knot nematodes. PhD Thesis, Dept of Crop and Soil Sciences, Univ of Georgia

Johnson HW, Bernard RL (1962) Soybean genetics and breeding. Adv Agron 14:149-221

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

Kwon SH, Torrie JH (1964) Heritability of and interrelationships among traits of two soybean populations. Crop Sci 4:196-198

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

Lee SH, Bailey MA, Mian MAR, Shipe ER, Ashley DA, Parrott WA, Hussey RS, Boerma HR (1996a) Identification of quantitative trait loci for plant height, lodging, and maturity in a soybean population segregating for growth habit. Theor Appl Genet 92:516-523

Lee SH, Bailey MA, Mian MAR, Carter TE, Jr, Shipe ER, Ashley DA, Parrott WA, Hussey RS, Boerma HR (1996b) RFLP loci associated with soybean seed protein and oil content across populations and locations. Theor Appl Genet 93:649-657

Li Z, Jakkula L, Hussey RS, Tamulonis JP, Boerma HR (2001) SSR mapping and confirmation of the QTL from PI96354 conditioning soybean resistance to southern root-knot nematode. Theor Appl Genet 103:1167-1173

Mansur LM, Orf JH, Chase K, Jarvik T, Cregan PB, Lark KG (1996) Genetic mapping of agronomic traits using recombinant inbred lines of soybean. Crop Sci 36:1327-1336

Mansur LM, Lark KG, Kross H, Oliveria 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 (1996) Molecular-marker analysis of seed weight: Genomic locations, gene action, and evidence for orthologous evolution among three legume species. Theor Appl Genet 93:574-579

Melchinger AE, Utz HF, Schön CC (1998) Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power

of QTL detection and large bias in estimates of QTL effects. Genetics 149:383-403

Mian MAR, Bailey MA, Tamulonis JP, Shipe ER, Carter TE, Jr, Parrott WA, Ashley DA, Hussey RS, Boerma HR (1996) Molecular markers associated with seed weight in two soybean populations. Theor Appl Genet 93:1011-1016

Miller JE, Fehr WR (1979) Direct and indirect recurrent selection for protein in soybeans. Crop Sci 4:196-198

Mudge J, Cregan PB, Kenworthy JP, Kenworthy WJ, Orf JH, Young ND (1997) Two microsatellite markers that flank the major soybean cyst nematode resistance locus. Crop Sci 37:1611-1615

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

Qiu BX, Arelli PR, Sleper DA (1999) RFLP markers associated with soybean cyst nematode resistance and seed composition in a 'Peking' x 'Essex' population. Theor Appl Genet 98:356-364

SAS Institute (1989) SAS/STAT User's Guide, Version 6, 4th edn, Vol 1. SAS Institute, Inc., Cary, North Carolina

Sebolt AM, Shoemaker RC, Diers BW (2000) Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. Crop Sci 40:1438-1444

Smith RR, Weber CR (1968) Mass selection by specific gravity for protein and oil soybean populations. Crop Sci 8:373-377

Tanksley SD, Young ND, Paterson AH, Bonierbale MW (1989) RFLP mapping in plant breeding: new tools for an old science. Biotechnology 7:257-264

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

Wilson LA (1999) Current developments in soyfood processing North American. World Soybean Res Conf VI Proc, Chicago, Illinois, pp 394-402

Young ND (1999) A cautiously optimistic vision for marker-assisted breeding. Molecular Breeding 5:505-510

Table 4.1. Previously reported (Lee et al. 1996b; Mian et al. 1996) independent RFLP markers associated with seed protein, seed oil, and seed weight in the soybean population of PI97100 x Coker 237 evaluated in this study.

Trait	Linkage Group	RFLP marker	R^{2b}	Allelic means[a]		
				P/P	P/C	C/C
			(%)	----------g/kg----------		
Seed protein	E	A454-1	8.8	449	445	442
	H	A566-2	13.5	451	445	443
	K	A065-1	10.6	442	444	449
	UNK	A132-4	13.3	441	446	448
				----------g/kg----------		
Seed oil	C1	A063-1	13.2	191	191	188
	G	L154-2	17.1	189	192	189
	H	A566-2	9.8	188	190	192
				--------mg/seed--------		
Seed weight	D2	A257-1	8.0	133	135	141
	G	A235-1	10.0	132	139	136
	M	Cr529-1	11.0	140	135	130

[a] P/P: homozygous for PI97100; P/C: heterozygous; C/C: homozygous for Coker 237.
[b] Percent of the phenotypic variation explained by the markers.

Table 4.2. RFLP loci associated with seed composition and seed weight in the independent soybean population of PI97100 x Coker 237 that consisted of 176 F_2-derived lines.

		Combined					Environment					
				Allelic means[c] ----g/kg----			Athens 1995		Athens 1996		Plains 1996	
Trait	LG/RFLP locus	*P*[a]	R^2(%)[b]	P/P	P/C	C/C	*P*[a]	R^2(%)[b]	*P*[a]	R^2(%)[b]	*P*[a]	R^2(%)[b]
Protein	E/A454-1	0.0000	12.3	457	450	446	0.0000	11.7	0.0001	10.8	0.0046	6.4
	H/A566-2	0.1319	-	451	450	447	0.1520	-	0.2799	-	0.2149	-
	K/A065-1	0.4199	-	449	451	449	-	-	-	-	-	-
	UNK/A132-4	0.0116	5.6	448	448	454	0.0690	3.4	0.0683	3.4	0.0070	6.2
Oil	C1/A063-1	0.0011	8.0	187	184	183	0.0352	4.0	0.0231	4.5	0.0003	9.7
	G/L154-2	0.9214	-	185	185	185	-	-	-	-	-	-
	H/A566-2	0.0008	8.3	185	184	187	0.0085	5.7	0.0149	5.0	0.0024	7.1
				----mg/seed----								
Seed weight	D2/A257-1	0.1696	-	159	162	165	0.3187	-	0.1364	-	0.2844	-
	G/A235-1	0.2384	-	158	162	163	0.5941	-	0.2484	-	0.0921	-
	M/Cr529-1	0.9261	-	161	161	162	0.7511	-	0.9947	-	0.8994	-

[a] Probability of a significant difference based on an F-test.
[b] Percent of the phenotypic variation explained by the marker.
[c] P/P: homozygous for PI97100; P/C: heterozygous; C/C: homozygous for Coker 237.

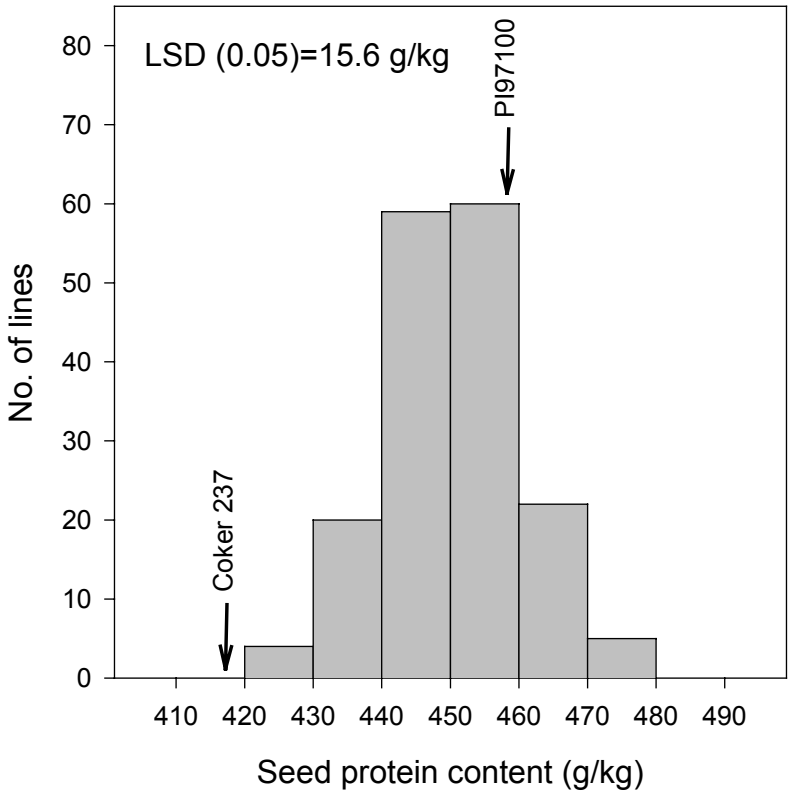

Figure 4.1. Frequency distribution for seed protein content of 180 F_2-derived lines of the soybean population PI97100 x Coker 237. The mean protein content of the population was 450 g/kg and the progeny exhibited a normal distribution.

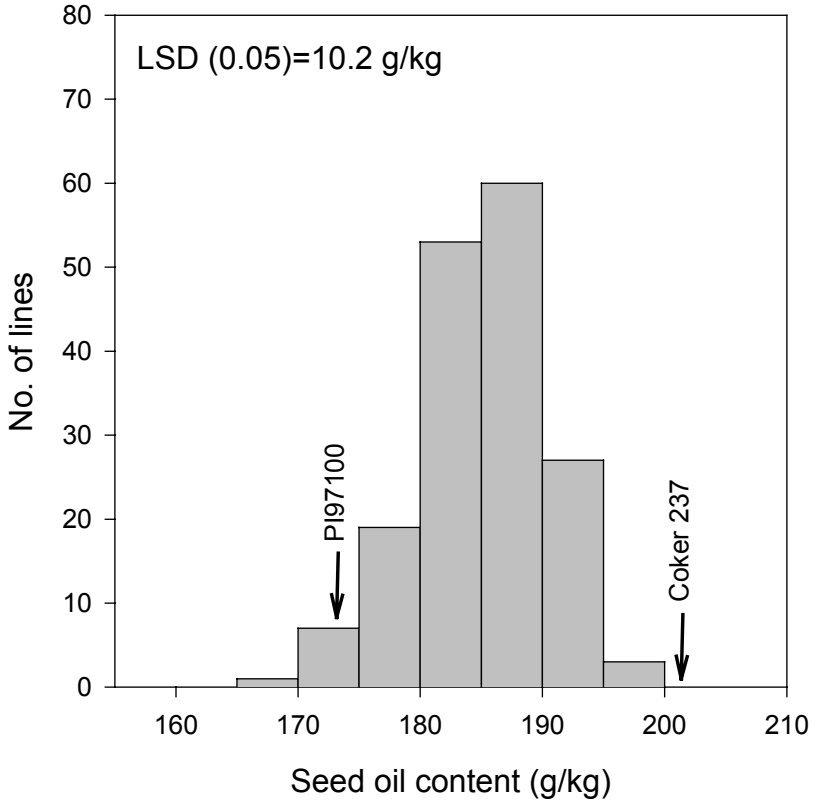

Figure 4.2. Frequency distribution for seed oil content of 180 F_2-derived lines of the soybean population PI97100 x Coker 237. The mean oil content of the population was 185 g/kg and the progeny exhibited a normal distribution.

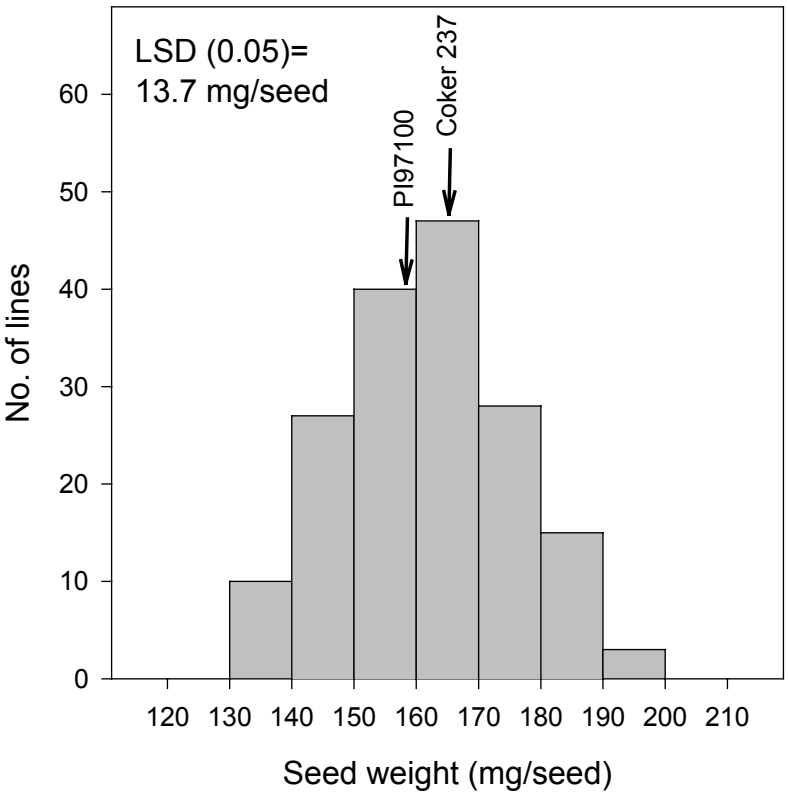

Figure 4.3. Normal frequency distribution for seed weight of 180 F_2-derived lines of the soybean population PI97100 x Coker 237. The mean seed weight of the population was 161 mg/seed and the progeny exhibited transgressive segregation.

CHAPTER 5

DIVERGENT HONEYCOMB SELECTION FOR SEED PROTEIN AND OIL CONTENT WITHIN SOYBEAN CULTIVARS[1]

[1]Fasoula, V.A. and H.R. Boerma. To be submitted to Crop Science.

Abstract

Soybean [*Glycine max* (L.) Merr.] seed is a major source of protein for animal feed and oil for human consumption. Elite soybean cultivars combine high productivity and pest resistance with broad adaptability over the target environments. Elite cultivars possess a limited amount of genetic variation, which is generally believed to be too small to make significant progress. Honeycomb breeding emphasizes selection after the release of cultivars to further improve the crop yield and quality of elite cultivars and to circumvent cultivar deterioration. The main objective of this study was to investigate the presence of genetic variation for seed composition traits within three elite soybean cultivars by honeycomb selection. Single plants from the three cultivars were grown in a replicated-3 honeycomb design using a plant-to-plant spacing of 90 cm. A total of 333 plants from Benning, 392 plants from Haskell, and 371 plants from Cook were evaluated. Divergent honeycomb selection for high and low protein and oil content was performed to select a total of 20 plants for protein and 20 plants for oil from each cultivar. The selected plants from the honeycomb experiment were evaluated in replicated row-plot experiments in subsequent years in order to (i) confirm the 1995 differences in row plots and (ii) select the most divergent lines with high and low protein and oil content. Our results indicate that honeycomb selection was successful in discovering a significant amount of genetic variation for seed composition within each of the three soybean cultivars. For protein content, the magnitude of variation between the highest- and lowest-protein lines averaged 5% (19 g kg^{-1}) across the three cultivars and ranged from 3 to 6 %. For oil composition, the magnitude of genetic variation between the most

divergent selections averaged 6% (12 g kg^{-1}) across the three cultivars and ranged from 5 to 7%. Significant variation was also discovered for fatty acid content although selection for the various fatty acids was not performed. The magnitude of variation averaged 11 g kg^{-1} for palmitic, 6 g kg^{-1} for stearic, 29 g kg^{-1} for oleic, 22 g kg^{-1} for linoleic, and 10 g kg^{-1} for linolenic acid across the three cultivars. The genetic variation discovered within the soybean cultivars is most likely due to latent variation, newly created variation, or epigenetic variation. Our results show that the exploitation of genetic variation within soybean cultivars will be very useful in improving the seed protein and oil content of elite cultivars.

Introduction

Soybean seed is a major source of protein for animal feed and oil for human consumption. It supplies approximately 65% of the world's protein meal and 25% of the world's edible oils (Golbitz, 2001). The high concentration of protein in soybean meal makes the meal a valuable livestock feed. Soybean protein has an excellent balance of amino acids compared with other vegetable proteins. In the United States, the annual production of soybean meal is approximately 34,686 Mg (Golbitz, 2001). More than 95% of this meal is used as animal feeds for poultry, swine, and cattle. For poultry, it has been determined that increasing the protein content of the soybean seed by 1% will increase the value of one Mg of soybean by approximately $4.00.

Soybean oil is an important source of vegetable oil for human food (Glaudemans et al., 1998). However, crude soybean oil has flavor and stability limitations that can be overcome only by expensive processing steps. The flavor problem of soybean oil has been attributed to the oxidation of fatty acids, primarily linolenic acid (Smouse, 1979). Soybean oil consists mainly of palmitic (C16:0), stearic (C18:0), oleic (C18:1), linoleic (C18:2), and linolenic (C18:3) acid. New soybean genotypes with modified fatty acid composition are constantly being considered for commercial production. A reduction in saturated fatty acids in soybean oil is desirable in order to reduce the human health risks associated with the consumption of saturated fats. Simultaneous increases in protein and oil content can proceed only to a limited extent, and most experimental data show that protein and oil content are negatively correlated (Johnson and Bernard, 1962; Kwon and Torrie, 1964; Smith and Weber, 1968; Burton, 1987).

As the world's major crops continue under intense selection, interest in genetic diversity is growing and has been described in many crops, including major field crops such as maize (*Zea mays* L.), soybean, and wheat (*Triticum aestivum* L.). Breeding programs in soybean and other crops usually restrict the use of parents to those improved for a variety of traits. Progress is still sufficient to encourage continued breeding within narrow genetic pools although each cycle is expected to reduce genetic variation. There is a lot of interest in genetic diversity, primarily due to the belief that progress will be hampered without adequate genetic variation. On the contrary, evidence from selection experiments within fairly homogeneous genetic pools suggests that the genome is more flexible and plastic than previously assumed. In her Nobel address, McClintock (1984) argued that the genome is dynamic and that it could modify itself under a variety of stress conditions.

Gordon and Byth (1972) reported significant variation within the predominantly self-pollinated tobacco (*Nicotiana tabacum* L.) cultivar Hicks. The authors found large differences for several agronomic traits among selections derived from Hicks and suggested that substantial advances can be made from selection among these strains. Byth and Weber (1968) found genetic variability within F_5-derived soybean lines that are considered to be relatively homozygous. In tomato (*Lycopersicon esculentum* L.), exploitable genetic variation was uncovered in advanced generations, after the point of achieving theoretical homozygosity (F_7 generation) (Christakis and Fasoulas, 2002).

Evidence from long-term selection experiments and doubled haploid studies also suggest that the genome is more flexible and plastic than previously assumed. The Univ. of Illinois long-term selection studies for modified oil and protein in maize are especially

interesting (Dudley and Lambert, 1992). These lines have been selected for more than 90 generations and variation is still enough to achieve progress from selection. Sprague et al. (1960) and Russell et al. (1963) reported that doubled haploid lines of maize soon accumulated considerable variation in agronomic traits. These lines should have been homozygous at every locus and the authors argued that the observed variation could not be accounted for by the commonly reported rates of mutation.

Furthermore, Rasmusson and Phillips (1997) reported that in barley (*Hordeum vulgare* L.) incremental genetic gains were made for several traits in what appears to be a very narrow gene pool. They emphasized that elite gene pools have inherent mechanisms to provide a continuing source of new genetic variation. They hypothesized that selection gain occurs due to variation present in the original gene pool as well as due to *de novo* generated variation. Classical and molecular genetic analyses have shown that many mechanisms exist to generate *de novo* variation, such as intragenic recombination, unequal crossing over, DNA methylation, transposable elements, and gene amplification (Rasmusson and Phillips, 1997).

Elite cultivars possess a limited amount of genetic variation, which is generally believed to be too small to make significant progress. Fasoula (1990) worked with a long-established wheat cultivar Siete Cerros released by CIMMYT in 1966 and applied divergent honeycomb selection for high and low yield to 2,700 plants of Siete Cerros. Successful divergent selection within the wheat cultivar identified lines that produced 8% higher and 9% lower yields than Siete Cerros when evaluated at commercial crop densities. In cotton (*Gossypium hirsutum* L.), honeycomb selection for high yield per plant and quality was applied to 10,000 plants of the elite cultivar Sindos 80 and led to

the development of the cultivar Macedonia. Macedonia exhibited a 10% average superiority over Sindos 80 across sixteen locations and two years. Honeycomb selection within Macedonia continued for two more years and the best lines were grown at two diverse sites, a *Verticillium*-infected and a *Verticillium*-free site. Honeycomb selection for high yield per plant was applied to both sites and led to the isolation of two lines that significantly outyielded the best check at both environments and were tolerant to *Verticillium* wilt, in which the original cultivar Sindos 80 was susceptible (Fasoulas, 2000).

Honeycomb breeding emphasizes selection after the release of cultivars to further improve the crop yield and quality of elite cultivars and to circumvent cultivar deterioration (Fasoula and Fasoula, 2000). Continuous selection after the release of cultivars is imposed by the need to eliminate deleterious mutations and exploit any positive source of existing and newly derived variation, either genetic or epigenetic. The most fundamental principle of honeycomb breeding is that selection and evaluation is performed on single plants grown in the absence of competition. The masking effect of competition on the efficiency of single-plant selection has been reviewed by Fasoula and Fasoula (1997). Honeycomb designs are utilized to sample effectively for environmental diversity and to select objectively the most superior entries as well as the best plants within the selected entries (Fasoulas and Fasoula, 1995).

No reported studies have investigated the presence of genetic variation within elite soybean cultivars. Successfully grown soybean cultivars combine high productivity and pest resistance with broad adaptability over the target environments. The release of cultivars is a time consuming task, thus it is important to exploit any positive source of

latent and newly created genetic variation. The objectives of this study were to (i) investigate the presence of genetic variation within elite soybean cultivars for seed composition traits by honeycomb selection, and (ii) investigate whether this variation can be exploited in a soybean breeding program to improve the protein and oil content of elite cultivars.

Materials and Methods

Three elite soybean cultivars (Benning, Haskell, and Cook) released by the Georgia Agric. Exp. Stns. and highly adapted to southern USA were chosen as the selection material. Haskell and Benning are Maturity Group VII cultivars and were released in 1993 and 1995, respectively (Boerma et al., 1994, 1997). Cook is classified as Maturity Group VIII and was released in 1991 (Boerma et al., 1992). Each cultivar was developed from a different cross. Benning was derived from a F_4 plant, whereas Haskell and Cook were derived from F_5 plants. The generations were advanced by the single pod-bulk method to the F_4 or F_5 generation in Georgia and Puerto Rico.

The three soybean cultivars were grown in a replicated-3 honeycomb design where the identification codes 1, 2, and 3 corresponded to plants from Benning, Haskell, and Cook, respectively (Fig. 1). The design samples effectively for spatial heterogeneity by allocating the plants of any cultivar in the corners of equilateral triangles and every plant in the center of a complete circular replicate (Fig. 1). The honeycomb trial was planted on 16 June 1995 in hill plots with 3 seeds per hill. The seed source of the cultivars was foundation seed produced in 1994. Three weeks after planting, each hill

plot was thinned to 1 plant per plot. Plants were grown in the absence of competition using a plant-to-plant spacing of 90 cm (1.4 plants/m^2) and a row spacing of 76 cm. Since the plants were grown in a noncompetitive environment no border plots were necessary.

A total of 333 plants from Benning, 392 plants from Haskell, and 371 plants from Cook were evaluated. At maturity, each hill was harvested by hand and threshed by a plot combine. Data were collected for seed quality, seed yield, seed protein, and seed oil content. Divergent honeycomb selection for high and low protein and oil content was based on plants with extreme seed composition values and acceptable seed quality. A total of 10 plants for high and 10 plants for low protein from each cultivar were selected. Similarly, a total of 10 plants for high and 10 plants for low oil from each cultivar were selected.

The selected plants from the honeycomb experiment were evaluated in replicated experiments in subsequent years in order to (i) confirm the 1995 differences in row plots and (ii) select the most divergent lines with high and low protein and oil content. Three field experiments were conducted in 1996, one for each cultivar, and each experiment included a total of 44 entries (40 selected plants from the honeycomb trial and four checks). The four checks in each experiment represented the original soybean cultivar from which selection was initiated. The three experiments were planted on 13 June 1996 at the Univ. of Georgia Plant Sciences Farm near Athens GA, in randomized complete block designs with three replications. Seeds from the selected single plants were planted in one-row plots with a row spacing of 76 cm and a row length of 3.5 m. At maturity, each plot was harvested and data on seed protein and seed oil content were collected.

Based on the 1996 analysis of data for seed protein and oil content, the most divergent lines in each group whose mean was statistically ($P < 0.05$) different from the mean of lines in the other group were evaluated in 1997. For each cultivar, the lines with high and low seed composition were grown in a randomized complete block experiment along with some other lines of the same cultivar selected for agronomic traits. Three separate field experiments were established, one for each cultivar. Each experiment consisted of 36 entries with three replications, including four entries of the original soybean cultivar as checks. The experiments were planted on 6 June 1997 at the Univ. of Georgia Plant Sciences Farm near Athens, GA and on 10 June 1997 at the Univ. of Georgia Southwest Branch Experiment Station near Plains, GA. At both locations, the experimental unit for each entry was two 4-m rows spaced 76 cm apart. At maturity, each plot was harvested and threshed with a plot combine. Data were collected on seed yield, seed protein, and seed oil content.

Based on the combined analysis across 1996 and 1997 for seed protein and oil content, the most divergent lines in each group that were statistically ($P < 0.05$) different from all lines in the other group were evaluated in 1998. More specifically, six lines were selected from Benning for protein (three high and three low) and six lines were selected for oil content (three high and three low). For Cook, ten lines were evaluated for seed protein (seven high and three low) and seven lines were evaluated for seed oil (four high and three low). For Haskell, eight lines were evaluated for protein content (five high and three low) and seven lines were evaluated for oil composition (three high and four low).

For each cultivar, the lines with high and low protein or oil content were grown in a randomized complete block design along with some other lines of the same cultivar

selected for agronomic traits. Three separate field experiments were established in 1998, one for each cultivar. Each experiment consisted of 30 entries with four replications, including four entries of the original soybean cultivar as checks. The experiments were planted on 1 June 1998 at the Univ. of Georgia Plant Sciences Farm and on 9 June 1998 at the Univ. of Georgia Southwest Branch Experiment Station. At both locations, the experimental unit for each entry was two 4-m rows spaced 76 cm apart. At maturity, each plot was harvested and threshed with a plot combine. Data were collected on seed yield, seed protein, seed oil, and seed fatty acid content.

For protein and oil content determination, a 50-g seed sample from each plot was sent to the USDA-ARS National Center for Agricultural Utilization Research at Peoria, IL for seed composition analysis. A 18- to 20-g sample of seed was analyzed for protein and oil composition with a model 1255 Infratec NIR food and feed grain analyzer. The protein and oil values were converted to a moisture-free basis. The seed fatty acid content was analyzed using gas-liquid chromatography of the methyl esters. The analyses were performed in the USDA-ARS, National Center for Agricultural Utilization Research at Peoria, IL (USDA-ARS/Peoria).

Data were analyzed by analysis of variance or nearest neighbor analysis when significant field trends were present (Stroup and Mulitze, 1991) using the Agrobase Software (Agronomix Software Inc., Canada). For comparing the mean of a line with the mean of the check, the least significant difference (LSD) was calculated using the equation $LSD = t_{df}\ 0.05\ [EMS\ (1/n1 + 1/n2)]^{1/2}$, where EMS = error term used for estimating the significance, n1 = number of values used in computing the mean of a line, and n2 = number of values used in calculating the mean of the check. For the statistical

model, replications, locations, and years were considered random effects and lines were considered a fixed effect.

Results

Honeycomb selection within soybean cultivars

The protein content of the 333 single plants of Benning averaged 403 g kg^{-1} and individual plants varied by 66 g kg^{-1} (Fig. 2). The individual Benning plants averaged 201 g kg^{-1} oil content and ranged from 180 to 217 g kg^{-1}. The distribution of protein content among 371 single plants of Cook ranged from 402 to 466 g kg^{-1} and averaged 435 g kg^{-1} (Fig. 3). For oil content, the individual Cook plants varied by 50 g kg^{-1} and averaged 186 g kg^{-1}. The 392 single plants of Haskell averaged 394 g kg^{-1} protein and 199 g kg^{-1} oil (Fig. 4). The individual Haskell plants with the extreme protein or oil contents differed by 88 g kg^{-1} in protein and 45 g kg^{-1} in oil content. The protein and oil contents of individual plants within each of the three cultivars appeared to be normally distributed (Figs. 2-4).

There were significant negative associations between oil and protein content among the entire population of single plants of Benning ($r = -0.54$, $P < 0.05$), Cook ($r = -0.75$, $P < 0.05$), and Haskell ($r = -0.62$, $P < 0.05$). This is in agreement with other studies that have reported negative correlations for protein and oil contents (Johnson and Bernard, 1962; Kwon and Torrie, 1964; Smith and Weber, 1968; Burton, 1987). Divergent honeycomb selection for the 10 plants with the highest and lowest protein and oil content within each of the three soybean cultivars was conducted. Seeds from the selected single plants were further evaluated in replicated row-plot experiments in order to advance the most divergent lines with high and low protein and oil content.

Genetic variation for seed composition within soybean cultivars

When averaged across the three years of replicated row-plot experiments (five environments), the three lines of Benning selected for high protein (H-4123, -2403, and -1935) produced significantly ($P < 0.05$) more protein than the three lines selected for low protein (L-2317, -4048, and -1818) (Table 1). The three high-protein lines averaged 11 g kg^{-1} more protein and 8 g kg^{-1} less oil content than the three low-protein lines. In addition, the three high-protein selections (H-4123, -2403, and -1935) were significantly ($P < 0.05$) higher in protein content (9, 7, and 7 g kg^{-1}) than Benning and two low-protein selections (L-4048 and -1818) were significantly ($P < 0.05$) lower (3 and 4 g kg^{-1}) than Benning (Table 1). Averaged across the three years, there was a difference of 13 g kg^{-1} in protein content between the most divergent lines H-4123 and L-1818.

The oil content of the three high-oil single plant selections from Benning (H-335, -4048, and -353) averaged 11 g kg^{-1} higher oil content than the three low-oil selections (L-1120, -2403, and -4123) (Table 2). Averaged across 3 years, each of the high-oil selections averaged more oil than Benning and each of the low-oil selections averaged less oil than Benning. The high-oil lines were consistently higher in oil content than the low-oil lines in each year (Table 2). The three high-oil selections averaged 7 g kg^{-1} less in protein content than the three low-oil lines. The highest-oil line H-335 averaged 213 g kg^{-1} and the lowest-oil line L-4123 averaged 199 g kg^{-1}. This represents a variation of 14 g kg^{-1} in oil composition within the cultivar Benning.

When averaged across 3 years of replicated row-plot experiments (five environments), the seven lines of Cook selected for high protein (H-1346, -3614, -3038, -3546, -3746, -2734, and -2014) produced significantly ($P < 0.05$) more protein than the three lines selected for low protein (L-3008, -2116, and -1035) (Table 3). The seven high-protein lines averaged 15 g kg^{-1} more protein content and 5 g kg^{-1} less oil content than the three low-protein lines. The high-protein lines were consistently higher in protein content than the low-protein lines across years and in each individual year. The seven high-protein lines were significantly ($P < 0.05$) higher in protein content (11, 10, 10, 9, 9, 8, and 6 g kg^{-1}) than Cook (Table 3). In addition, two low-protein selections (L-2116 and -1035) were significantly ($P < 0.05$) lower (7 and 8 g kg^{-1}) in protein content than Cook. Across 3 years, there was a significant difference of 19 g kg^{-1} in protein content between the most divergent lines H-1346 and L-1035 derived from Cook.

The oil content of the four high-oil lines selected from within Cook (H-2116, -1035, -4014, and -3402) was 7 g kg^{-1} higher than the three low-oil lines (L-2734, -1346, and -3746) when averaged over 3 years (Table 4). The four high-oil lines averaged 12 g kg^{-1} less protein content than the three low-oil lines. Furthermore, the four high-oil lines of Cook were significantly ($P < 0.05$) higher in oil composition (4, 4, 3, and 2 g kg^{-1}) than Cook and the three low-oil lines had significantly ($P < 0.05$) lower oil (3, 4, and 5 g kg^{-1}) than Cook. Averaged across 3 years, the difference in oil composition between the extreme lines H-2116 and L-3746 was 9 g kg^{-1} (Table 4).

Averaged across 3 years (five environments), the five lines of Haskell selected for high protein (H-425, -515, -551, -4243, and -1452) produced significantly ($P < 0.05$) more protein than the three low-protein selections (L-1736, -2646, and -1112) (Table 5).

The five high-protein lines averaged 14 g kg^{-1} more protein content and 8 g kg^{-1} less oil content than the three low-protein lines. Furthermore, the five high-protein lines had significantly ($P < 0.05$) higher protein content (18, 13, 9, 8, and 4 g kg^{-1}) than Haskell, whereas the three low-protein lines had significantly ($P < 0.05$) lower protein (3, 3, and 6 g kg^{-1}) than Haskell (Table 5). Averaged across 3 years, there was a significant difference of 24 g kg^{-1} in protein content between the most divergent lines H-425 and L-1112 derived from Haskell.

The three lines selected from within Haskell for high oil were significantly ($P < 0.05$) different from the four lines selected for low oil when averaged across 3 years (Table 6). The three high-oil lines selected from Haskell (H-2646, -2151, and -339) averaged 10 g kg^{-1} more in oil content and 10 g kg^{-1} less in protein content than the four low-oil lines (L-2303, -425, -4243, and -515). The high-oil lines were consistently higher in oil composition than the low-oil lines across years and in each individual year. H-2646 line produced significantly ($P < 0.05$) higher oil content (3 g kg^{-1}) than Haskell and the four low-oil lines produced significantly ($P < 0.05$) lower oil content (5, 7, 9, and 10 g kg^{-1}) than Haskell (Table 6). Averaged across 3 years, the difference in oil composition between the most divergent lines H-2646 and L-515 was 13 g kg^{-1}.

Genetic variation for fatty acids within soybean cultivars

For each cultivar, there were 26 lines derived from honeycomb selection that were evaluated for seed composition and agronomic traits at two locations in 1998. Although these lines were not selected for fatty acid content, there were significant ($P < 0.05$)

differences among the lines for all the fatty acids. The fatty acid distribution of the 26 lines derived from Benning for palmitic (A), oleic (B), linoleic (C), and linolenic acid (D) is shown in Fig. 5. A significant difference among lines of 5, 25, 19, and 11 g kg^{-1} for palmitic, oleic, linoleic, and linolenic acid, respectively, was found. The lines with the highest and the lowest fatty acid content were significantly ($P < 0.05$) different from the original Benning. For the cultivar Cook, significant ($P < 0.05$) differences of 17 g kg^{-1} for palmitic acid, 24 g kg^{-1} for oleic acid, 19 g kg^{-1} for linoleic acid, and 7 g kg^{-1} for linolenic acid were found (Fig. 6). In addition, the lines with the highest and the lowest fatty acid content were significantly ($P < 0.05$) different from the original Cook.

The 26 lines derived from Haskell exhibited a greater amount of variation for fatty acid content than either Benning or Cook (Fig. 7). For palmitic and stearic acid, a significant ($P < 0.05$) difference of 12 g kg^{-1} and 10 g kg^{-1} was found, respectively. Consistent with the other cultivars, the largest amount of variation was observed for oleic acid. The lines with the lowest and the highest oleic acid content differed by 37 g kg^{-1} (Fig. 7). For linoleic acid, a significant difference of 29 g kg^{-1} was found. The lines with the highest and the lowest fatty acid content were significantly different ($P < 0.05$) from the original Haskell.

Discussion

Our results show that honeycomb selection was successful in discovering a significant amount of genetic variation for seed protein and oil within each of the three soybean cultivars. For protein content, the magnitude of genetic variation between the

highest- and lowest-protein lines averaged 5% (19 g kg^{-1}) across the three cultivars and ranged from 3 to 6%. For oil composition, the magnitude of variation between the highest- and lowest-oil selections averaged 6% (12 g kg^{-1}) across the three cultivars and ranged from 5 to 7%. Although selection for the various fatty acids was not in our initial plan, significant variation was discovered for fatty acid content among the lines selected for seed protein, oil, or other agronomic traits. Across the three cultivars, the magnitude of within-cultivar variation averaged 9% (11 g kg^{-1}) for palmitic acid, 16% (6 g kg^{-1}) for stearic acid, 13% (29 g kg^{-1}) for oleic acid, 5% (22 g kg^{-1}) for linoleic acid, and 14% (10 g kg^{-1}) for linolenic acid. These results show that the genetic diversity discovered within cultivars can be further exploited in a breeding program to improve the seed composition traits of elite cultivars.

In soybean, inverse relationships between seed protein and seed oil content have been reported (Kwon and Torrie, 1964; Smith and Weber, 1968; Burton, 1987). In our results, a negative correlation between protein and oil composition was observed in some of the selected lines. Some lines though, did not exhibit this relationship. For example, lines H-335 and H-353 selected from Benning had 6 g kg^{-1} and 4 g kg^{-1} higher oil content than the original Benning, but the same protein content as Benning (Table 2). Among the lines derived from Cook, four high-protein lines (H-3614, -3038, -3546, and -2014) had 10, 10, 9, and 6 g kg^{-1} higher protein than the original Cook, but the same oil content as Cook (Table 3). In addition, two high-oil lines (H-4014 and -3402) had the same protein content as Cook, but they produced 3 g kg^{-1} and 2 g kg^{-1} higher seed oil than Cook (Table 4). Among the lines selected from within Haskell, two high-protein lines (H-551 and -1452) had 9 g kg^{-1} and 4 g kg^{-1} higher protein than the original Haskell, but the same seed

oil as Haskell (Table 5). Two high-oil lines (H-2151 and -339) had the same protein content with Haskell, but 2 g kg^{-1} higher oil content (Table 6).

In addition, inverse relationships between seed protein content and seed yield have been reported in soybean (Burton, 1987; Wilcox and Cavins, 1995). This relationship has limited the progress in the development of high-protein and high-yielding soybean cultivars. In our study, a negative correlation between seed protein content and seed yield was not observed in any of the selected lines (Tables 1, 3, 5). For Benning and Cook, all the selected lines for high and low seed composition had the same yield as the original cultivar (Tables 1-4). Among the lines derived from Haskell, two high-protein lines (H-4243 and -1452) produced 8 g kg^{-1} and 4 g kg^{-1} higher protein content than Haskell. Furthermore, these lines exhibited 21% (3388 kg ha^{-1}) and 14% (3175 kg ha^{-1}) higher seed yield ($P < 0.05$) than the original Haskell (Table 5). Thus, a positive association between seed protein and seed yield was exhibited for these lines. A low-oil line of Haskell (L-2303) produced 5 g kg^{-1} lower oil and 12% (2451 kg ha^{-1}) significantly ($P < 0.05$) lower seed yield than Haskell (Table 6). Lines like H-4243 and L-2303 indicate the importance of selection within elite soybean cultivars.

Elite cultivars have long been considered by plant breeders a relatively permanent record with non-existing or very limited genetic variation. But increasingly, molecular biologists are finding that the genome undergoes constant remodeling and restructuring. Our results corroborate those of Rasmusson and Phillips (1997) who reported that in barley, incremental genetic gains were made for several traits in what appears to be a very narrow gene pool. They hypothesized that selection gain occurs due to variation present in the original gene pool as well as due to *de novo* generated variation. *De novo*

sources include intragenic recombination, unequal crossing over, DNA methylation, transposable elements, and gene amplification.

Evidence from long-term selection experiments and doubled haploid studies (Sprague et al., 1960; Russell et al., 1963; Dudley and Lambert, 1992) also suggest that the genome is more flexible and plastic than previously assumed. In tomato, exploitable genetic variation was uncovered in advanced generations, after the point of achieving theoretical homozygosity (F_7 generation) (Christakis and Fasoulas, 2002). A comparative RFLP analysis of sunflower (*Helianthus annuus*) cultivars and an evaluation of cultivar variation in rice (*Oryza sativa* L.), using microsatellite and RFLP markers, revealed the existence of within-cultivar variation (Zhang et al., 1995; Olufowote et al., 1997).

Honeycomb breeding emphasizes selection after the release of cultivars to further improve the crop yield and quality of elite cultivars and to circumvent cultivar deterioration. In wheat, divergent honeycomb selection within the cultivar Siete Cerros produced lines with 8% higher and 9% lower yield than Siete Cerros (Fasoula, 1990). Honeycomb selection within the elite cotton cultivar Sindos 80 led to the development of the cultivar Macedonia, which exhibited a 10% average superiority over Sindos 80 across sixteen locations and two years. Honeycomb selection within Macedonia continued for two more years and the best lines were grown at two diverse sites, a *Verticillium*-infected and a *Verticillium*-free site. Honeycomb selection for high yield per plant was applied to both sites and led to the isolation of two lines that significantly outyielded the best check at both environments and were tolerant to *Verticillium* wilt, in which the original cultivar Sindos 80 was susceptible (Fasoulas, 2000).

In our study, the variation discovered within the soybean cultivars is most likely due to latent variation within F_4- or F_5- derived lines, newly-created variation, and/or epigenetic variation in response to environmental stimuli. McClintock (1984) has argued that the genome is dynamic and that it can modify itself under a variety of stress conditions. Our results show that the exploitation of genetic variation within soybean cultivars will be very useful in improving the seed protein and oil content of elite cultivars. In addition, they provide evidence that the genome is not a permanent record, but a dynamic and flexible entity.

References

Boerma, H.R., R.S. Hussey, D.V. Phillips, E.D. Wood, and S.L. Finnetry. 1992. Registration of 'Cook' soybean. Crop Sci. 32:497.

Boerma, H.R., R.S. Hussey, D.V. Phillips, E.D. Wood, and S.L. Finnetry. 1994. Registration of 'Haskell' soybean. Crop Sci. 34:541.

Boerma, H.R., R.S. Hussey, D.V. Phillips, E.D. Wood, G.B. Rowan, and S.L. Finnetry. 1997. Registration of 'Benning' soybean. Crop Sci. 37:1982.

Burton, J.W. 1987. Quantitative genetics: Results relevant to soybean breeding. In: Wilcox, J.R. (ed) Soybeans: Improvement, production and uses. 2nd ed. Agron. Monogr. 16. ASA, CSSA, and SSSA, Madison, WI.

Byth. D.E., and C.R. Weber. 1968. Effects of genetic heterogeneity within two soybean populations I. Variability within environments and stability across environments. Crop Sci. 8:44-47.

Christakis, P.A., and A.C. Fasoulas. 2002. The effects of the genotype by environmental interaction on the fixation of heterosis in tomato. J. Agric. Sci. 138:1-5.

Dudley, J.W., and R.J. Lambert. 1992. Ninety generations of selection for oil and protein in maize. Maydica 37:81-87.

Fasoula, D.A. 1990. Correlations between auto-, allo- and nil-competition and their implications in plant breeding. Euphytica 50:57-62.

Fasoula, D.A., and V.A. Fasoula. 1997. Competitive ability and plant breeding. Plant Breed. Rev. 14:89-138.

Fasoula, V.A., and D.A. Fasoula. 2000. Honeycomb breeding: principles and applications. Plant Breed. Rev. 18:177-250.

Fasoulas, A.C. 2000. Building up resistance to *Verticillium* wilt in cotton through honeycomb breeding. p. 120-124. In: F.M. Gillham (ed.) New frontiers in cotton research. Proc. 2nd World Cotton Research Conference. Sept. 6-12, 1998. Athens, Greece.

Fasoulas, A.C., and V.A. Fasoula. 1995. Honeycomb selection designs. Plant Breed. Rev. 13:87-139.

Glaudemans, H.D., M.M.J. Timmermans, and H. Rijkse. 1998. The world of edible oils. p. 1-5. Rabobank International Marketing, Utrecht, the Netherlands.

Golbitz, P. 2001. 2001 Soya & Oilseed Bluebook. 2001. Soyatech, Inc., Bar Habor, ME.

Gordon, I.L., and D.E. Byth. 1972. Comparisons among strains of the tobacco cultivar Hicks illustrating variability within a single cultivar. Queensland J. Agricultural and Animal Sci. 29:255-264.

Johnson, H.W., and R.L. Bernard. 1962. Soybean genetic and breeding. Adv. Agron. 14:149-221.

Kwon, S.H., and J.H. Torrie. 1964. Heritability and interrelationships among traits of two soybean populations. Crop Sci. 4:196-198.

McClintock, B. 1984. The significance of the responses of the genome to challenge. Science 226:792-801.

Olufowote, J.O., Y. Xu, X. Chen, W.D. Park, H.M. Beachell, R.H. Dilday, M.Goto, and S.R. McCouch. 1997. Comparative evaluation of within-cultivar variation of rice (*Oryza sativa* L.) using microsatellite and RFLP markers. Genome 40:370-378.

Rasmusson, D.C., and R.L. Phillips. 1997. Plant breeding progress and genetic diversity from de novo variation and elevated epistasis. Crop Sci. 37:303-310.

Russell, W.A., G.F. Sprague, and L.H. Penny. 1963. Mutations affecting quantitative characters in long-time inbred lines of maize. Crop Sci. 3:175-178.

Smith, R.R., and C.R. Weber. 1968. Mass selection by specific gravity for protein and oil soybean populations. Crop Sci. 8:373-375.

Smouse, T.H. 1979. A review of soybean oil reversion flavor. J. Am. Oil Chem. Soc., 56:747A-751A.

Sprague, G.F., W.A. Russell, and L.H. Penny. 1960. Mutations affecting quantitative traits in selfed progeny of doubled monoploid maize stocks. Genetics 45:855-865.

Stroup, W.W., and D.K. Mulitze. 1991. Nearest neighbor adjusted best linear unbiased prediction. Am. Stat. 45:194-200.

Wilcox, J.R., and J.F. Cavins, 1995. Backcrossing high seed protein to a soybean cultivar. Crop Sci. 35:1036-1041.

Zhang, Y.X., L. Gentzbittel, F. Vear, and P. Nicolas. 1995. Assessment of inter- and intra-inbred line variability in sunflower (*Helianthus annuus*) by RFLPs. Genome 38:1040-1048.

Table 5.1. Genetic variation for seed protein content within the soybean cultivar Benning. Honeycomb selection in 1995 and subsequent evaluations in 1996-1998 identified three lines with high protein and three lines with low protein content, designated with the letter H and L, respectively.

Lines from honeycomb selection	Seed protein content				Mean‡ oil content	Mean‡ seed yield
	1996	1997†	1998†	Mean‡ (% Benning)		
	g kg^{-1}					kg ha^{-1}
H-4123	415	423	418	419 (102)	199	2969
H-2403	412	421	415	417 (102)	200	3148
H-1935	413	422	413	417 (102)	204	2843
L-2317	408	411	405	408 (100)	207	3304
L-4048	403	410	407	407 (99)	213	3016
L-1818	400	412	404	406 (99)	208	2907
LSD (0.05)	8	5	5	4	3	NS
Benning	406	415	408	410 (100)	207	2859
LSD (0.05)§	7	4	4	3	3	NS

† Means in 1997 and 1998 are from two locations (Athens GA and Plains GA).
‡ Means for protein and oil were combined over five environments (3 years). Means for seed yield were combined over 2 years.
§ Least significant difference for comparisons of the Benning mean to the mean of a selected line.

Table 5.2. Genetic variation for seed oil content within the soybean cultivar Benning. Honeycomb selection in 1995 and subsequent evaluations in 1996-1998 identified three lines with high oil and three lines with low oil content, designated with the letter H and L, respectively.

Lines from honeycomb selection	Seed oil content				Mean‡ protein content	Mean‡ seed yield
	1996	1997†	1998†	Mean‡ (% Benning)		
	g kg^{-1}					kg ha^{-1}
H-335	232	201	214	213 (103)	412	3008
H-4048	229	205	212	213 (103)	407	3016
H-353	229	202	212	211 (102)	411	2741
L-1120	221	193	203	203 (98)	414	3009
L-2403	214	190	204	200 (97)	417	3148
L-4123	218	186	201	199 (96)	419	2969
LSD (0.05)	7	5	3	3	4	NS
Benning	223	198	209	207 (100)	410	2859
LSD (0.05)§	5	4	3	3	3	NS

† Means in 1997 and 1998 are from two locations (Athens GA and Plains GA).
‡ Means for protein and oil were combined over five environments (3 years). Means for seed yield were combined over 2 years.
§ Least significant difference for comparisons of the Benning mean to the mean of a selected line.

Table 5.3. Genetic variation for seed protein content within the soybean cultivar Cook. Honeycomb selection in 1995 and subsequent evaluations in 1996-1998 identified seven lines with high protein and three lines with low protein content, designated with the letter H and L, respectively.

Lines from honeycomb selection	Seed protein content				Mean‡ oil content	Mean‡ seed yield
	1996	1997†	1998†	Mean‡ (% Cook)		
	g kg^{-1}					kg ha^{-1}
H-1346	441	427	430	431 (103)	193	2953
H-3614	448	426	425	430 (102)	196	2899
H-3038	446	422	429	430 (102)	196	2983
H-3546	440	425	428	429 (102)	197	2896
H-3746	441	426	426	429 (102)	192	3055
H-2734	441	424	426	428 (102)	194	2866
H-2014	442	423	422	426 (101)	198	2966
L-3008	428	414	415	417 (99)	199	2964
L-2116	424	409	413	413 (98)	201	2862
L-1035	417	405	416	412 (98)	201	2804
LSD (0.05)	12	5	5	5	3	NS
Cook	427	417	420	420 (100)	197	2918
LSD (0.05)§	11	4	4	4	2	NS

† Means in 1997 and 1998 are from two locations (Athens GA and Plains GA).
‡ Means for protein and oil were combined over five environments (3 years). Means for seed yield were combined over 2 years.
§ Least significant difference for comparisons of the Cook mean to the mean of a selected line.

Table 5.4. Genetic variation for seed oil content within the soybean cultivar Cook. Honeycomb selection in 1995 and subsequent evaluations in 1996-1998 identified four lines with high oil and three lines with low oil content, designated with the letter H and L, respectively.

Lines from honeycomb selection	Seed oil content				Mean‡ protein content	Mean‡ seed yield
	1996	1997†	1998†	Mean‡ (% Cook)		
	g kg^{-1}					kg ha^{-1}
H-2116	212	196	201	201 (102)	413	2862
H-1035	209	197	200	201 (102)	412	2804
H-4014	212	197	198	200 (102)	420	2807
H-3402	210	195	198	199 (101)	422	3016
L-2734	205	188	195	194 (98)	428	2866
L-1346	205	188	193	193 (98)	430	2953
L-3746	204	185	193	192 (97)	429	3055
LSD (0.05)	5	5	3	3	5	NS
Cook	207	192	197	197 (100)	420	2918
LSD (0.05)§	4	4	2	2	4	NS

† Means in 1997 and 1998 are from two locations (Athens GA and Plains GA).
‡ Means for protein and oil were combined over five environments (3 years). Means for seed yield were combined over 2 years.
§ Least significant difference for comparisons of the Cook mean to the mean of a selected line.

Table 5.5. Genetic variation for seed protein content within the soybean cultivar Haskell. Honeycomb selection in 1995 and subsequent evaluations in 1996-1998 identified five lines with high protein and three lines with low protein content, designated with the letter H and L, respectively.

Lines from honeycomb selection	Seed protein content				Mean‡ oil content	Mean‡ seed yield
	1996	1997†	1998†	Mean‡ (% Haskell)		
	g kg^{-1}					kg ha^{-1}
H-425	425	423	425	424 (104)	197	3011
H-515	418	415	424	419 (103)	195	2913
H-551	415	413	416	415 (102)	204	2861
H-4243	422	409	414	414 (102)	196	3388
H-1452	413	405	412	410 (101)	204	3175
L-1736	406	399	406	403 (99)	206	2939
L-2646	397	403	406	403 (99)	208	2951
L-1112	401	396	404	400 (98)	206	2958
LSD (0.05)	9	6	5	4	4	414
Haskell	404	404	408	406 (100)	205	2786
LSD (0.05)§	7	5	4	3	3	320

† Means in 1997 and 1998 are from two locations (Athens GA and Plains GA).
‡ Means for protein and oil were combined over five environments (3 years). Means for seed yield were combined over 2 years.
§ Least significant difference for comparisons of the Haskell mean to the mean of a selected line.

Table 5.6. Genetic variation for seed oil content within the soybean cultivar Haskell. Honeycomb selection in 1995 and subsequent evaluations in 1996-1998 identified three lines with high oil and four lines with low oil content, designated with the letter H and L, respectively.

Lines from honeycomb selection	Seed oil content				Mean‡ protein content	Mean‡ seed yield
	1996	1997†	1998†	Mean‡ (% Haskell)		
	g kg^{-1}					kg ha^{-1}
H-2646	223	204	203	208 (101)	403	2951
H-2151	225	202	204	207 (101)	405	2827
H-339	220	202	205	207 (101)	406	2909
L-2303	209	197	198	200 (98)	403	2451
L-425	210	193	196	198 (97)	424	3011
L-4243	202	190	198	196 (96)	414	3388
L-515	211	189	193	195 (95)	419	2913
LSD (0.05)	7	4	3	4	4	414
Haskell	217	203	202	205 (100)	406	2786
LSD (0.05)§	6	3	2	3	3	320

† Means in 1997 and 1998 are from two locations (Athens GA and Plains GA).
‡ Means for protein and oil were combined over five environments (3 years). Means for seed yield were combined over 2 years.
§ Least significant difference for comparisons of the Haskell mean to the mean of a selected line.

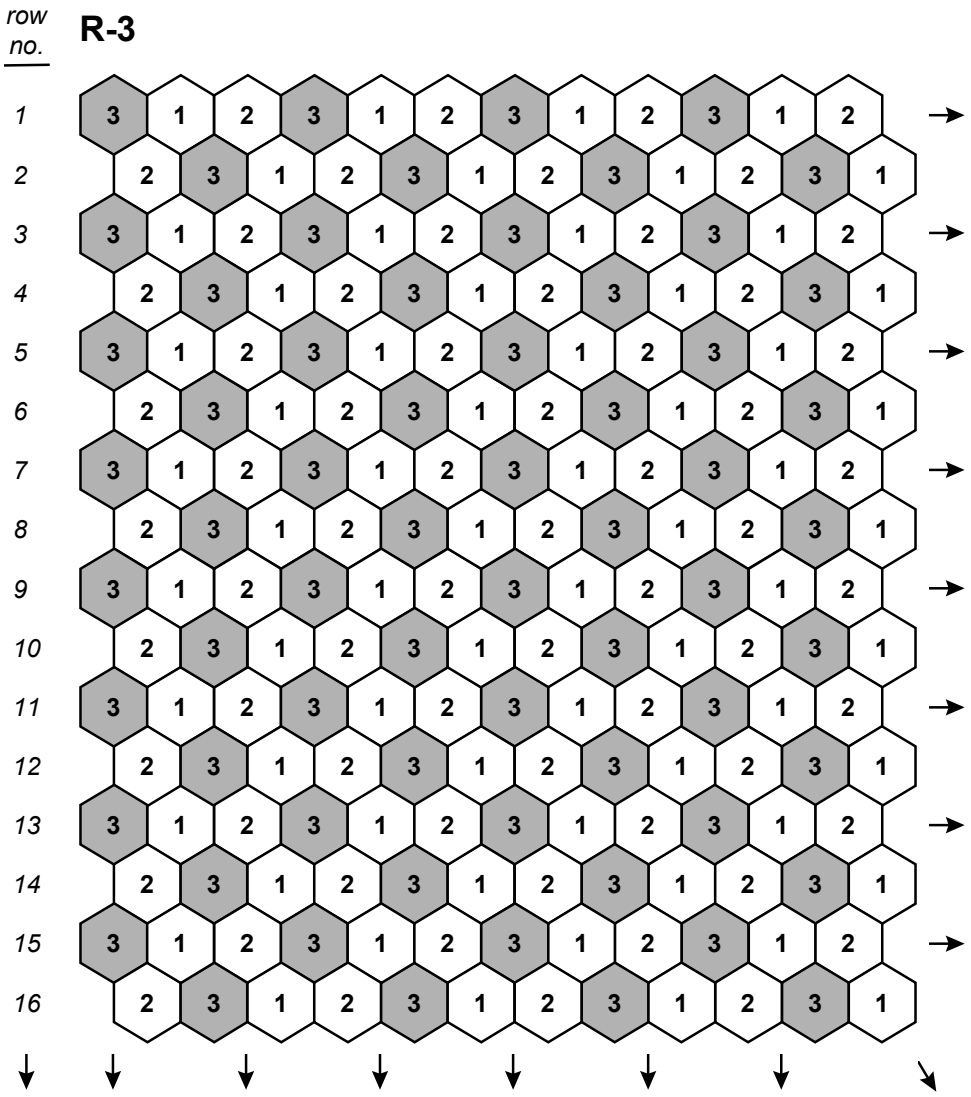

Figure 5.1. The replicated-3 honeycomb design evaluated plants from three cultivars, arranged in ascending order and in horizontal field rows, using a plant spacing of 90 cm (1.4 plants/m^2). The numbers in the figure represent the position of the plants in the field and the hexagons represent the area exploited by each plant. The identification codes 1, 2, and 3 corresponded to plants from Benning, Haskell, and Cook, respectively. Plants of any cultivar form an equilateral triangular lattice pattern (ETL) that samples effectively for spatial heterogeneity by allocating them evenly across the experimental site. The ETL pattern is exemplified for the single plants of Cook (gray hexagons).

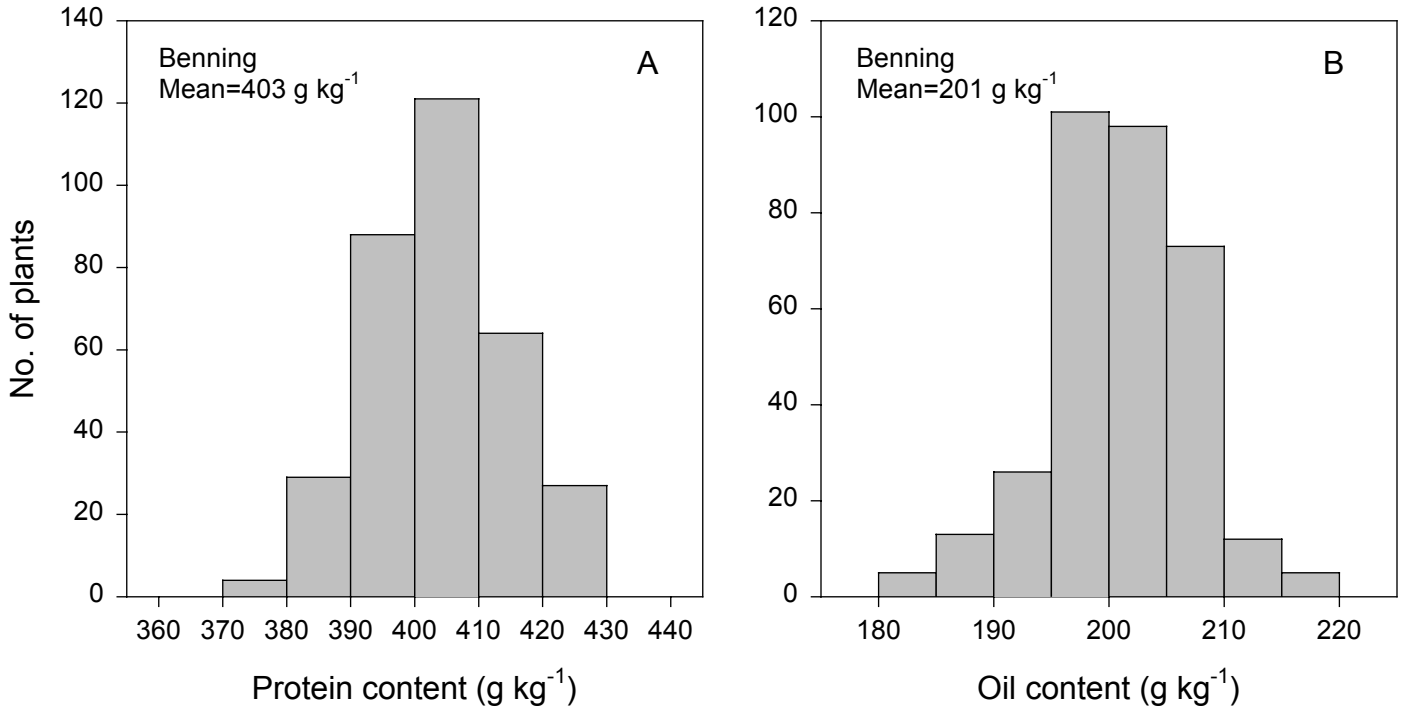

Figure 5.2. Frequency distribution of single plants (1.4 plants/m^2) from Benning for seed protein (A) and seed oil content (B).

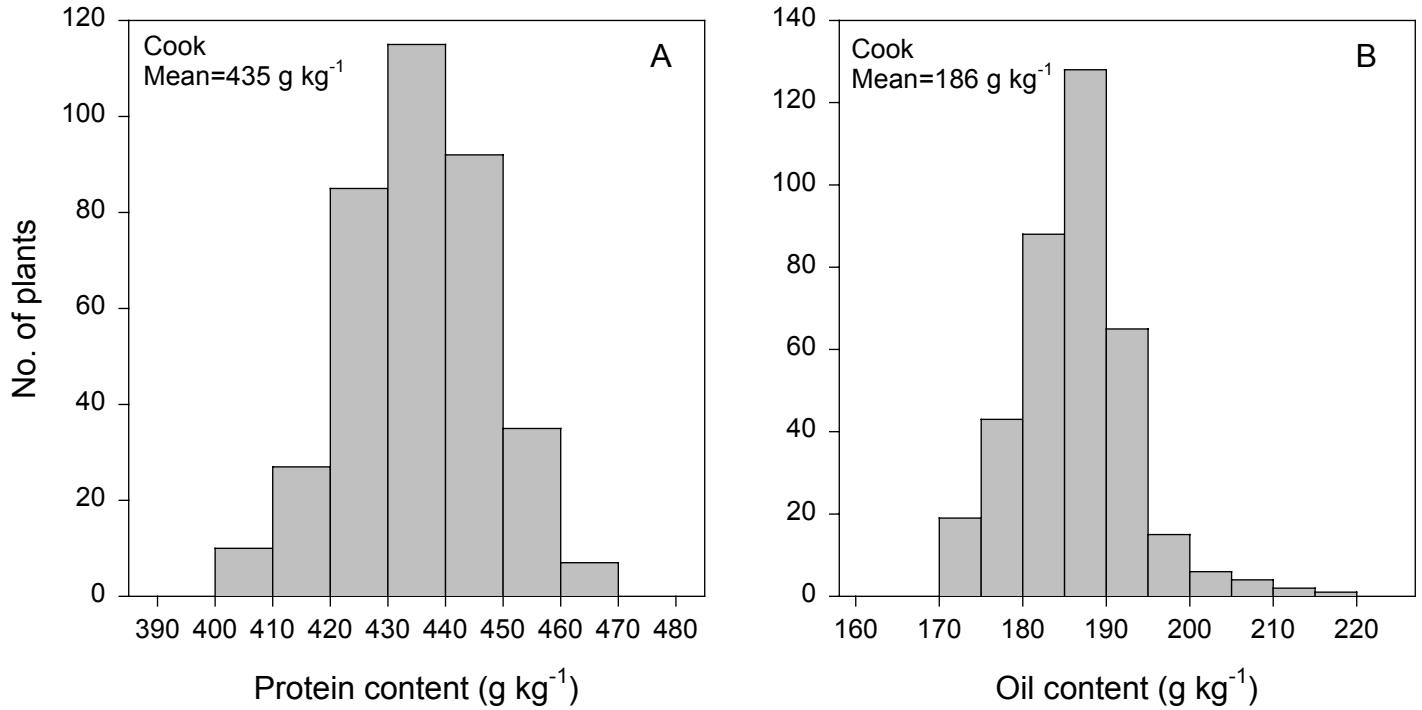

Figure 5.3. Frequency distribution of single plants (1.4 plants/m^2) from Cook for seed protein (A) and seed oil content (B).

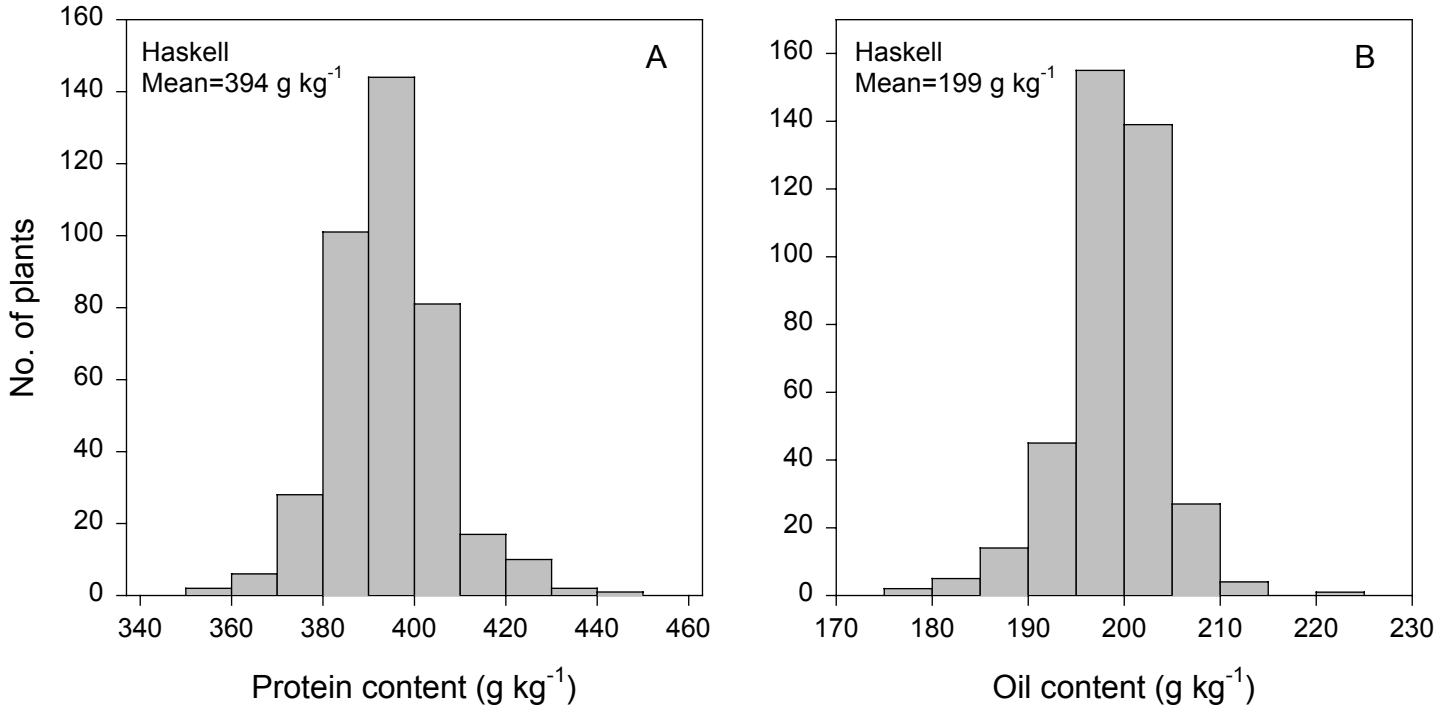

Figure 5.4. Frequency distribution of single plants (1.4 plants/m^2) from Haskell for seed protein (A) and seed oil content (B).

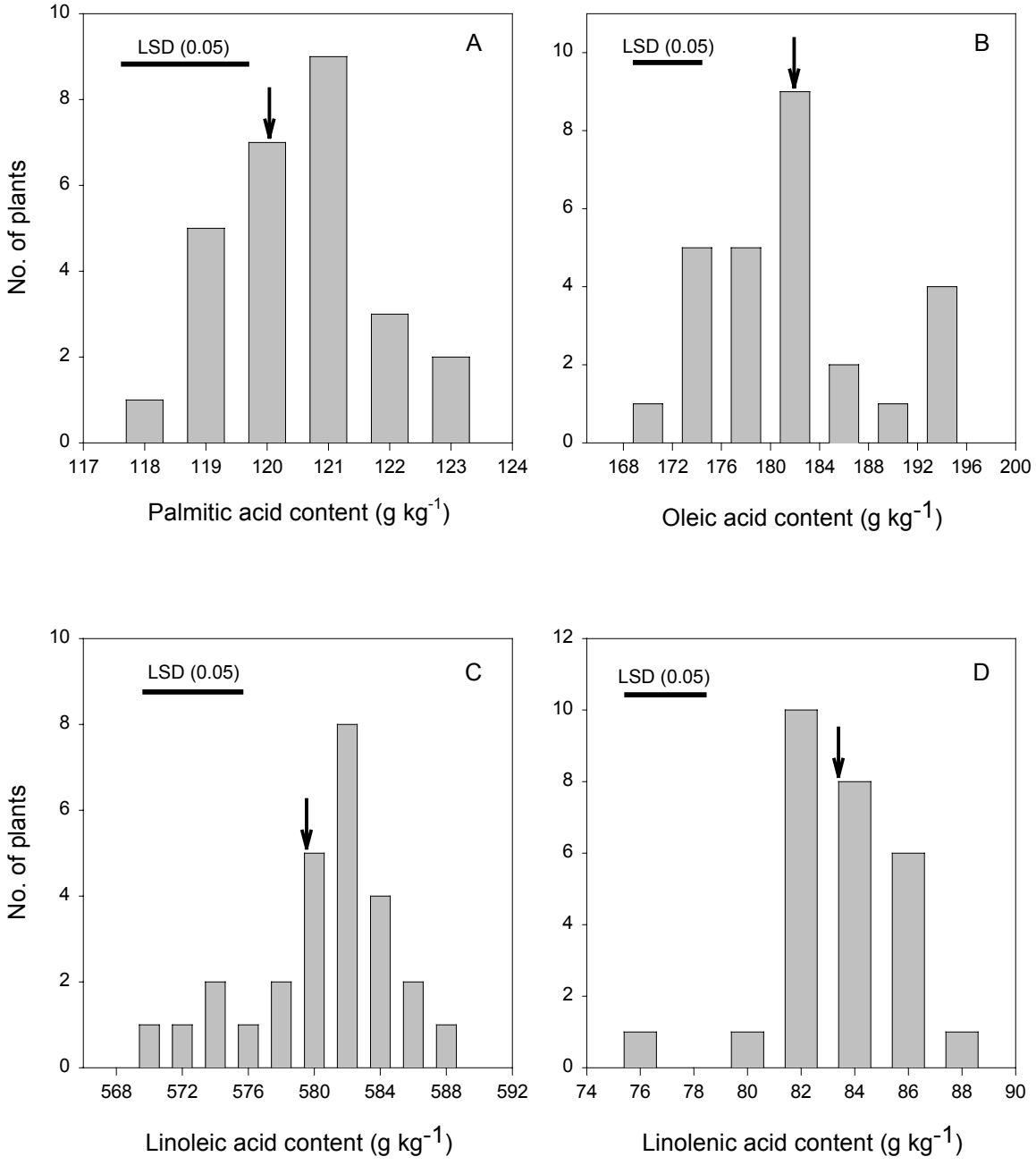

Figure 5.5. Genotypic variation among 26 lines derived from Benning for palmitic acid (A), oleic acid (B), linoleic acid (C), and linolenic acid (D). The arrow represents the value of the original Benning.

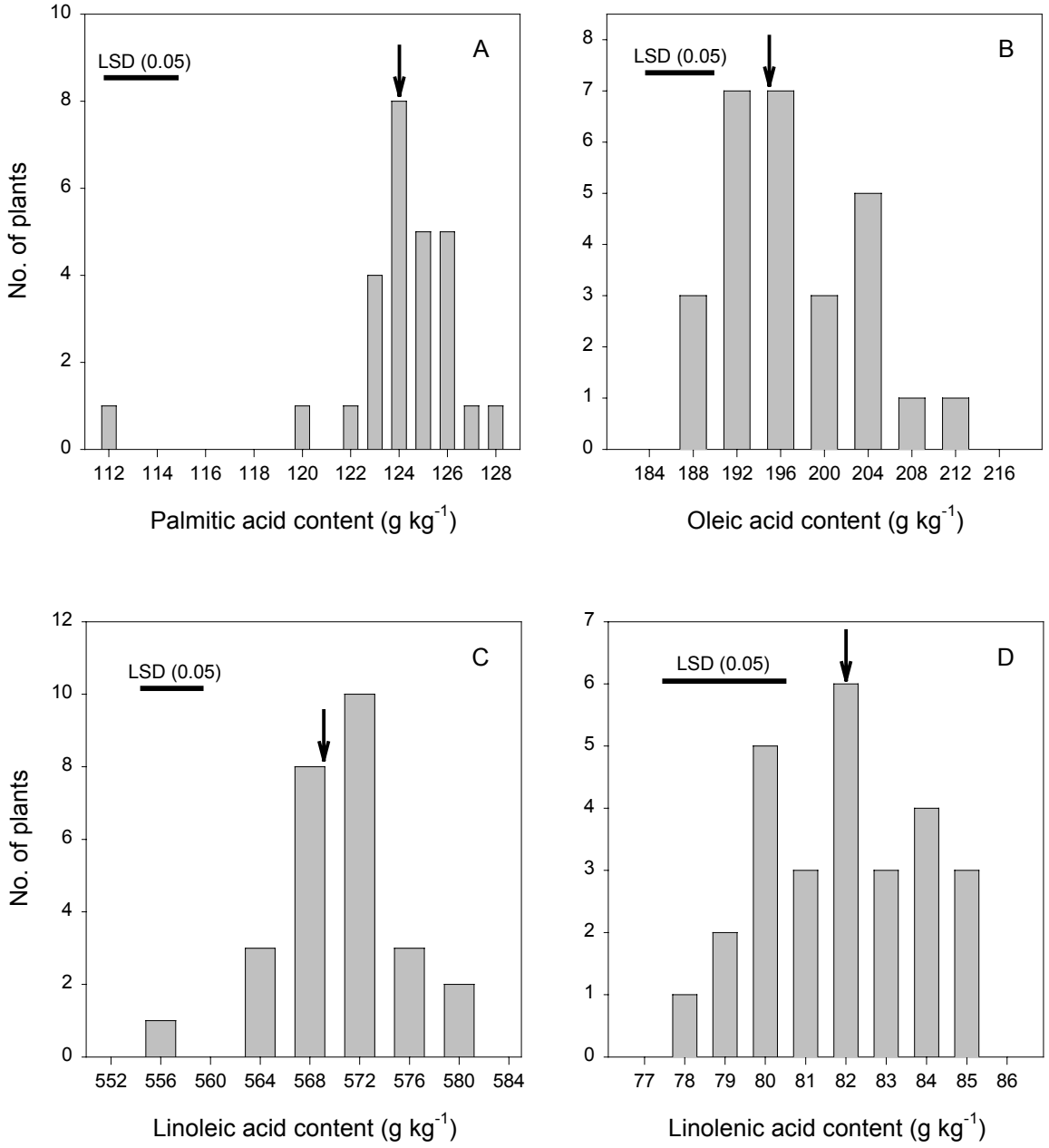

Figure 5.6. Genotypic variation among 26 lines derived from Cook for palmitic acid (A), oleic acid (B), linoleic acid (C), and linolenic acid (D). The arrow represents the value of the original Cook.

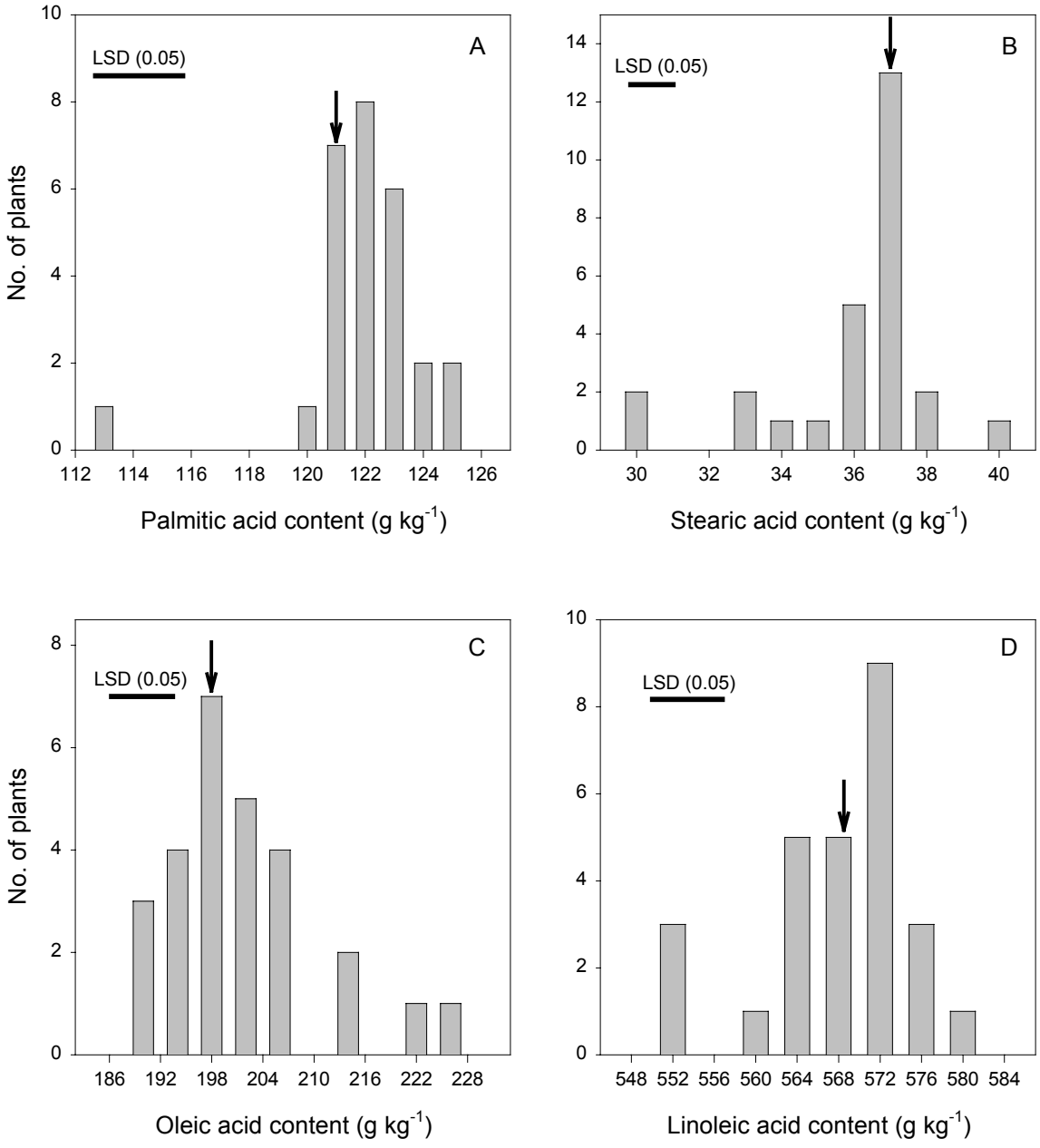

Figure 5.7. Genotypic variation among 26 lines derived from Haskell for palmitic acid (A), stearic acid (B), oleic acid (C), and linoleic acid (D). The arrow represents the value of the original Haskell.

CHAPTER 6

SUMMARY AND CONCLUSIONS

Three independent, but related studies were conducted to further the objective of improving soybean seed traits and resistance to tobacco ringspot virus. The objective of the first study was to identify and map the gene(s) for bud blight caused by tobacco ringspot virus. Bud blight can cause 25 to 100% yield reduction in soybean. Two populations derived from the cross of Young x PI416937 were evaluated. Young is resistant to bud blight and PI416937 is susceptible. One population consisted of 116 F_4-derived lines and was used to identify and map restriction fragment length polymorphism (RFLP) markers associated with resistance to bud blight. In the field, the plots were naturally infected with tobacco ringspot virus and were visually scored according to the number of plants that exhibited terminal bud death.

A major gene was identified and mapped by K644_1 on linkage group (LG) F. It accounted for 82% of the variation in bud blight score. Composite interval mapping located the single gene for bud blight resistance within the 8.2 cM interval between the RFLP markers K644_1 and A069_b. To verify the genomic location of the major bud blight gene, a second population of Young x PI416937 that consisted of 180 F_2-derived lines was evaluated. In this population, simple sequence repeat (SSR) markers were utilized. The major gene conditioning bud blight resistance was found to be closely

linked to the Satt510 marker, which is approximately 2 to 4 cM from RFLP marker K644_1 based on the USDA/Iowa State Univ. soybean genetic map.

The technological advancements in molecular biology have created the opportunity for genetic dissection of complex traits into a set of discrete loci. RFLP markers have been used extensively to map the genomic location of soybean quantitative trait loci (QTL) for many agronomic, physiological, and seed composition traits. In soybean and other crops, there is limited and imprecise information confirming the previously reported QTL. The objective of the second study was to utilize an independent F_2-derived soybean population of PI97100 x Coker 237 to confirm previously reported RFLP markers associated with seed protein, seed oil, and seed weight, mapped in an F_2-derived population created from the same parents. Single-factor analysis of variance (ANOVA) was used to confirm the RFLP loci that are significantly ($P < 0.05$) associated with seed composition and seed weight.

Two out of four previously reported QTL for seed protein, two out of three QTL for seed oil, and one out of three QTL for seed weight were confirmed in the independent population. Therefore, 50% of the QTL detected in the original mapping population were verified in the new population of PI97100 x Coker 237. The confirmed QTL were detected based on the mean phenotypic data across three environments as well as within each environment. These QTL were consistent across environments and were confirmed across two independent populations. The QTL that were not confirmed in this study may have been erroneously declared significant in the original population (Type I error) or they may have been specific for the sample of lines or environments used in the original population. These results confirm the necessity of mapping quantitative trait loci in

multiple environments and parallel populations before utilizing them in a plant improvement program.

Soybean seed is a major source of protein for animal feed and oil for human consumption. Elite soybean cultivars possess a limited amount of genetic variation, which is generally believed to be too small to make significant progress. The objective of the third study was to investigate the presence of genetic variation for seed composition traits within three elite soybean cultivars by honeycomb selection. Single plants from the three cultivars were grown in a replicated-3 honeycomb design using a plant spacing of 90 cm. Divergent honeycomb selection for seed protein and oil was performed to select a total of 20 plants for protein and 20 plants for oil from each cultivar. The selected plants from the honeycomb experiment were evaluated in replicated row-plot experiments for three years in order to select the most divergent lines with high and low protein or oil content.

The results indicated that honeycomb selection was successful in discovering a significant amount of genetic variation for seed composition within each of the three soybean cultivars. For protein content, the magnitude of variation between the highest- and lowest-protein lines averaged 5% (19 g kg^{-1}) across the three cultivars and ranged from 3 to 6 %. For oil composition, the magnitude of genetic variation between the most divergent selections averaged 6% (12 g kg^{-1}) across the three cultivars and ranged from 5 to 7%. Significant variation was also discovered for fatty acid content although selection for the various fatty acids was not performed. The magnitude of within-cultivar variation averaged 11 g kg^{-1} for palmitic, 6 g kg^{-1} for stearic, 29 g kg^{-1} for oleic, 22 g kg^{-1} for linoleic, and 10 g kg^{-1} for linolenic acid across the three cultivars. The genetic variation

discovered within the soybean cultivars is most likely due to latent variation, newly created variation, or epigenetic variation. These results show that the exploitation of the within-cultivar variation will be very useful in improving the seed protein and oil of elite soybean cultivars. In addition, they provide further evidence that the genome is not a permanent record but a dynamic and flexible entity.