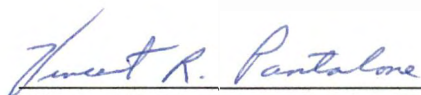


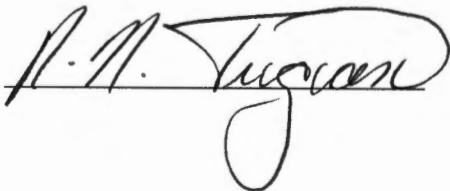
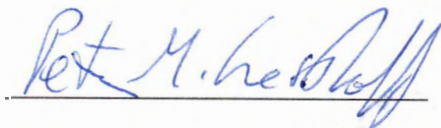
To the Graduate Council:

I am submitting herewith a thesis written by Angie Bell Chapman entitled "QTL and Heritability analysis of Essex x Williams F₂ and F_{4:6} Soybean Progenies." I have examined the final copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Plant and Soil Sciences.

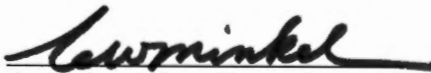


Vincent R. Pantalone, Major Professor

We have read this thesis
and recommend its acceptance:



Accepted for the Council:



Associate Vice Chancellor and
Dean of the Graduate School

**QTL AND HERITABILITY ANALYSIS OF
ESSEX x WILLIAMS
F₂ AND F_{4:6}
SOYBEAN PROGENIES**

A Thesis
Presented for the
Master of Science Degree
The University of Tennessee, Knoxville

Angie Bell Chapman
May 2000

AG-VET-MED.

Thesis
2000
.C389

ACKNOWLEDGMENTS

There are many people to whom I owe thanks. First I would like to thank Dr. Pantalone, for his assistance, patience, and thoughtfulness throughout this project. I would like to thank Dr. Allen, Dr. Trigiano, and Dr. Gresshoff for serving on my committee as well as Dr. Trigiano's lab assistant, Melissa Ament, and Dr. Gresshoff's assistant Janice Crockett. I would like to thank Artem Men, Jiri Stiller, Ali Ustun, and Madeleine Spencer for their technical assistance and conversation as well as everyone else in the Plant Molecular Genetics lab. I also wish to express thanks to Dr. Sams for his insight in helping me choose the correct course of study and Dr. Saxton for his help in understanding SAS and all other components of statistics.

I would particularly like to thank Debbie L. Ellis and fellow graduate student Caroline Cherrak as well as everyone else in the Soybean Breeding and Genetics Program: Beth Meyer, Jimmy Bice, Meredith Crosby, Cindy Longmire, and Amanda Pierce. I also owe thanks to Lee Ellis and all the guys at the Plant Science Farm for their invaluable help in the soybean fields.

Finally, I owe the greatest thanks to my generous parents, Avis and Don, my beautiful sister, Amy and my wonderful, fun, and loving husband Brian. I would also like to thank Meme, Grandmother, my recently deceased Grandfather for their support because they never had this opportunity and my late best friend Heather Rushing for her enduring friendship and optimism through thick and thin

ABSTRACT

The use of molecular markers in plant breeding for indirect selection of agronomically important traits could have a large impact on the efficiency of plant breeding systems. Most agronomically important traits are polygenic (quantitative) traits and the improvement of these traits through marker assisted breeding would have the greatest impact because traditional plant breeding for quantitative traits is usually slow and difficult.

The purpose of this research is to 1) identify quantitative trait loci (QTL) for the quantitative traits yield, plant height, seed protein concentration, seed oil concentration, seed size, and maturity, in an F_2 population from a cross between the soybean cultivars 'Essex' and 'Williams', 2) conduct a field test on $F_{4:6}$ pure lines to assess heritability and genetic variance for each quantitative trait, and 3) compare simple sequence repeat (SSR) markers which were found to be significant QTL's in the F_2 population will also be analyzed with field data from $F_{4:6}$ lines to identify whether early generation (F_2) molecular marker selection is effective and consistent with measured traits from more inbred ($F_{4:6}$) lines.

DNA was extracted from young F_2 generation soybean leaves and amplified via polymerase chain reaction (PCR) using SSR markers. The amplified products were then run on polyacrylamide gels and silver stained to visualize bands. The $F_{2:3}$ seeds were sent to Costa Rica for three generations of advancement during the winter of 1998-1999. The $F_{4:6}$ seeds were grown in Knoxville and Springfield, TN in summer of 1999. Markers that were polymorphic between the parents were then analyzed against the F_2 and $F_{4:6}$

generation trait data using the general linear model procedure of SAS. Additive and dominant effects were partitioned by utilizing a general QTL model and a maturity covariate model to detect QTL's. The genetic variance and heritability were calculated for each trait in the $F_{4:6}$ generation, over two environments, using the mixed model procedure of SAS.

For the F_2 population, significant additive QTL's were Satt540 (maturity), Satt52 (height), Satt239 (yield; seed size), Satt373 (oil; seed size; height), and significant dominant QTL's were Satt540 (maturity; yield; seed size), Satt52 (maturity; protein; oil), Satt373 (height), Sctt9 (height). However, soybean is cultivated as a pure-line crop. Hence, additive QTL's in later generations, and which are significant over environments, may provide the greatest utility to breeders. In the $F_{4:6}$ generation, averaged over both environments, the following significant additive QTL's were identified: Satt50 (yield, $r^2 = 1.8\%$), Satt251 (protein, $r^2 = 2.6\%$), Satt353 (seed size, $r^2 = 2.5\%$), Satt14 (seed size, $r^2 = 11.3\%$), and Satt239 (height, $r^2 = 2.6\%$).

Molecular breeding is becoming more practical as better technology emerges. Great progress has been made in the construction of genetic maps in many plant species, including all major crops. Most of these maps are constructed with RFLP markers, however many new maps are being constructed with both RADP and microsatellite markers. Continued molecular genetics and genomics research with our Essex x Williams population may further help to develop applied, molecular marker-assisted approaches for the benefit of soybean breeders.

TABLE OF CONTENTS

Section	Page
1. General Introduction	1
2. Literature Review	6
2.1 Quantitative Trait Loci (QTL)	6
2.2 Simple Sequence Repeat (SSR) Markers	13
2.3 Heritability and Genetic Variance	20
3. Introduction	22
3.1 Molecular Marker Study	22
3.2 Field Breeding Study	24
4. Materials and Methods	28
4.1 DNA Isolation and Characterization	28
4.2 DNA Quantification	30
4.3 Polymerase Chain Reaction (PCR)	30
4.4 Polyacrylamide Gel Electrophoresis (PAGE)	31
4.5 Silver Staining	32
4.6 Scoring	33
4.7 Field Breeding Study	33
5. Results and Discussion	36
5.1 F ₂ Population QTL Analysis	36
5.2 F _{4:6} Population QTL Analysis	53
5.3 Genetic Variance and Heritability	84

6. Concluding Remarks and Further Research	91
Literature Cited	93
Appendix	98
Vita	118

LIST OF TABLES

Table		Page
1	Mean and F ₂ Progeny Range of Maturity, Yield, Seed Protein, and Seed Oil.	37
2	QTL Analysis of Essex x Williams F ₂ Soybean Population for Plant Maturity at Knoxville, TN.	42
3	QTL Analysis of Essex x Williams F ₂ Soybean Population for Plant Yield at Knoxville, TN.	43
4	QTL Analysis of Essex x Williams F ₂ Soybean Population for Seed Protein Concentration at Knoxville, TN.	45
5	QTL Analysis of Essex x Williams F ₂ Soybean Population for Seed Oil Concentration at Knoxville, TN.	46
6	QTL Analysis of Essex x Williams F ₂ Soybean Population for Seed Size at Knoxville, TN.	48
7	QTL Analysis of Essex x Williams F ₂ Soybean Population for Plant Height at Knoxville, TN.	50
8	QTL Analysis of Essex x Williams F _{4:6} Soybean Population for Plant Maturity at Knoxville, TN.	54
9	QTL Analysis of Essex x Williams F _{4:6} Soybean Population for Plant Yield at Knoxville, TN.	56
10	QTL Analysis of Essex x Williams F _{4:6} Soybean Population for Seed Protein Concentration at Knoxville, TN.	57
11	QTL Analysis of Essex x Williams F _{4:6} Soybean Population for Seed Oil Concentration at Knoxville, TN.	59
12	QTL Analysis of Essex x Williams F _{4:6} Soybean Population for Seed Size at Knoxville, TN.	61
13	QTL Analysis of Essex x Williams F _{4:6} Soybean Population for Plant Height at Knoxville, TN.	63

14	QTL Analysis of Essex x Williams F _{4:6} Soybean Population for Plant Lodging at Knoxville, TN.	64
15	QTL Analysis of Essex x Williams F _{4:6} Soybean Population for Plant Yield at Highland Rim Experiment Station.	67
16	QTL Analysis of Essex x Williams F _{4:6} Soybean Population for Seed Protein Concentration at Highland Rim Experiment Station.	68
17	QTL Analysis of Essex x Williams F _{4:6} Soybean Population for Seed Oil Concentration at Highland Rim Experiment Station.	69
18	QTL Analysis of Essex x Williams F _{4:6} Soybean Population for Seed Size at Highland Rim Experiment Station.	70
19	QTL Analysis of Essex x Williams F _{4:6} Soybean Population for Plant Height at Highland Rim Experiment Station.	72
20	QTL Analysis of Essex x Williams F _{4:6} Soybean Population for Plant Lodging at Highland Rim Experiment Station.	73
21	QTL Analysis of Essex x Williams F _{4:6} Soybean Population for Plant Yield Combined over Knoxville and Springfield, TN.	75
22	QTL Analysis of Essex x Williams F _{4:6} Soybean Population for Seed Protein Concentration Combined over Knoxville and Springfield, TN.	76
23	QTL Analysis of Essex x Williams F _{4:6} Soybean Population for Seed Oil Concentration Combined over Knoxville and Springfield, TN.	78
24	QTL Analysis of Essex x Williams F _{4:6} Soybean Population for Seed Size Combined over Knoxville and Springfield, TN.	79
25	QTL Analysis of Essex x Williams F _{4:6} Soybean Population for Plant Height Combined over Knoxville and Springfield, TN.	80
26	QTL Analysis of Essex x Williams F _{4:6} Soybean Population for Plant Lodging Combined over Knoxville and Springfield, TN.	81
27	Genetic Variance, Heritability, Mean, and Progeny Range of the F _{4:6} Population Traits	85

A.1	Partial List of Soybean Cultivars with Essex x Williams in their lineage	99
A.2	Cultivar Descriptions of Essex and Williams	104
A.3	List of Screened Markers Showing Fragment Size for Markers that are Polymorphic Between Cultivars Essex and Williams.	105
A.4	Forward and Reverse Primer Sequences of Polymorphic SSR Markers.	107
A.5	Expected Mean Squares for Evaluating a Random Set of Genotypes over Random Environments.	108
A.6	Summarized List of Traits per Marker for the F ₂ Population at Knoxville, TN.	109
A.7	Summarized List of Traits per Marker for the F _{4:6} Population at Knoxville, TN.	111
A.8	Summarized List of Traits per Marker for the F _{4:6} Population at Highland Rim Experiment Station.	113
A.9	Summarized List of Traits per Marker for the F _{4:6} Population Combined over Knoxville and Springfield, TN.	115
A.10	Positive and Negative Parental Alleles, Linkage Groups, Base Pairs, and Trait Means for QTL's for an F _{4:6} Essex x Williams Population Combined over Knoxville and Springfield, TN.	117

LIST OF FIGURES

Figure		Page
1	Frequency Distribution for Plant Yield of Essex x Williams F ₂ Population.	38
2	Frequency Distribution for Seed Protein Concentration of Essex x Williams F ₂ Population.	39
3	Frequency Distribution for Mean Plot Yield of Essex x Williams F _{4:6} Population over Two Locations	87
4	Frequency Distribution for Mean Seed Protein Concentration of Essex x Williams F _{4:6} Population over Two Locations	88

1. GENERAL INTRODUCTION

Emerging from a relatively little known immigrant plant, the soybean, *Glycine max* (L.) Merrill, became a major grain crop in the United States. The soybean was cultivated largely for forage before 1940, but today it is the third largest in production of grain after *Zea mays* and *Triticum aestivum*, and second in value after *Zea mays*. The United States produces 51% of the world's yield with the Corn Belt and the lower Mississippi Delta being the major areas of production. Brazil, the People's Republic of China, and Argentina produce another 38%, while the remaining percentage is scattered among various countries in Asia and South America (Poehlman 1995).

Cultivated soybean has many uses in food and in industrial products. Soybean oil is the most widely used edible oil, accounting for 80% of the market for all edible fats and oils in the U.S.A. Industrial soybean oil products include surfactants, paints and coatings, solvents, lubricants, inks and biodiesel fuel. Examples of soy protein food products are, tofu, soy flour, and soymilk. Soy protein is used in industry to make products such as adhesives, plastics, and building materials such as particleboard and plywood (www.omnitechintl.com/soy_program/overview.html). Soybean meal is widely used in livestock feed, with poultry feed (52%) and swine feed (29%) predominating the soy meal market. The other 19% of the market is used in beef and dairy feed, pet food, human food and industrial products (www.soymeal.org/ksmith1.html).

Evidence suggests that soybean originated in the eastern half of northern China around the eleventh century BC with migration to central and southern China as well as

Korea by the first century AD. From the first century AD to about the 15th or 16th century, the Age of Discovery, soybeans were introduced into several countries with landraces developing in Japan, Indonesia, the Philippines, Vietnam, Thailand, Malaysia, Burma, Nepal, and northern India (www.hort.purdue.edu/newcrop/proceedings1990/V1-159.html).

Early introduction of the soybean to North America was by Samuel Bowen in 1765. Bowen was a former seaman employed by the East India Company. He brought soybeans to Savanna, Georgia from China via London, England. Bowen used his soybeans to manufacture soy sauce and vermicelli. In 1770 Benjamin Franklin grew soybeans on his farm in Philadelphia. In 1829 Professor Thomas Nuttall grew soybeans in the Botanic Gardens in Massachusetts. In 1851 the soybean was first introduced to Illinois and subsequently throughout the cornbelt. By 1854 soybean seeds were distributed to farmers throughout the United States (www.hort.purdue.edu/newcrop/proceedings1990/V1-159.html).

In 1949, Martin G. Weiss along with Jackson L. Carter, of the U.S. Regional Soybean Laboratory at Urbana, Illinois, initiated the development of a comprehensive soybean germplasm collection. Since 1954, Richard L. Bernard had been the curator of the northern collection in Urbana, IL. Edgar E. Hartwig was the curator of the southern soybean germplasm collection, located in Stoneville, Mississippi, from 1951 until his death in 1997. Today the USDA soybean collection contains about 13,000 strains of soybeans, wild soybeans, and wild perennial *Glycine* species (www.hort.purdue.edu/newcrop/proceedings1990/V1-159.html) and is currently managed by the national curator, Randall Nelson at Champaign, IL.

Soybean breeding in the U.S. developed as a cooperative between the United States Department of Agriculture and Agricultural Experiment Stations in the midwestern and southern states. At the University of Illinois, Champaign/Urbana in 1936, intensive research began on the soybean as a result of a cooperative federal/state research program, currently called the *U.S. Regional Soybean Laboratory*. Private seed companies were motivated to initiate soybean breeding programs as a result of the United States Plant Variety Protection Act of 1970, and today, research by private seed companies now exceeds that of publicly supported programs in the U.S.A. (Poehlman 1995).

Important breeding objectives for soybean are yield, maturity for the area of production, resistance to lodging and shattering, tolerance to environmental stresses, disease and nematode resistance, insect resistance, and seed quality and composition. Work is also being done in the nitrogen fixing aspect of the soybean (Poehlman 1995). Most of these traits are polygenic, or quantitative, traits which require years of breeding and selection.

The use of molecular markers in plant breeding for indirect selection of agronomically important traits could have a large impact on the efficiency of plant breeding systems. Some factors of plant breeding that could be improved with the help of molecular markers include: the identification and elimination of undesirable individuals in the early stages of selection, and identifying promising individuals prior to flowering to more accurately measure the expression of quantitative traits, and to facilitate the selection of several traits simultaneously (Allen 1994).

Most agronomically important traits are quantitative traits and the improvement of these traits through marker assisted breeding would have the greatest impact because

traditional plant breeding for quantitative traits is usually slow and difficult. Marker assisted selection can also greatly improve selection for traits of low heritability by using markers with high heritability. Selection of traits by genotype can be especially beneficial when the genotype x environment interaction is significant therefore masking the phenotype to be used in direct selection. It is thought that indirect selection of simply inherited traits is not needed because the phenotypes can be easily scored, but markers closely linked to qualitative traits such as pest and disease resistance could prove to be very valuable when considering quarantine laws against handling pests and diseases (Grattapaglia 1994). Furthermore, marker assisted backcross breeding can facilitate transfer of resistance genes while maintaining the desirable genetic background of the recurrent parent.

The purpose of this research is to 1) identify quantitative trait loci of quantitative traits consisting of, yield, plant height, protein concentration, oil concentration, seed size, and maturity, in an F_2 population from a cross between the soybean cultivars 'Essex' and 'Williams', 2) conduct a field test on $F_{4:6}$ pure lines to assess heritability, and genetic variance, and 3) simple sequence repeats which were found to be significant QTLs in the F_2 population will also be analyzed with field data from $F_{4:6}$ lines to identify whether early generation (F_2) molecular marker selection is effective and consistent with measured traits from more inbred ($F_{4:6}$) lines.

The parents Williams and Essex were chosen because they contribute prominently to modern northern and southern US soybean cultivars, respectively (Carter *et al.*, 1993). Table A.1 shows a list of cultivars with an Essex x Williams cross in their lineage. Essex and Williams are different for several qualitative and quantitative traits (Table A.2), and

have different ancestors, hence we would expect wide phenotypic diversity among progeny from a cross between these two parents.

2. LITERATURE REVIEW

2.1 QUANTITATIVE TRAIT LOCI (QTL)

Shoemaker *et al.*, (1994) reports that approximately forty to sixty percent of the soybean genome consists of repetitive sequences and that 65-70 % of single sequence repeats alternate copies of 1.1 - 1.4 kb single copy sequences with 0.3 -0.4 kb copies of repetitive sequences. Low or single copy clones are used to accurately analyze band segregation patterns. Restriction enzymes have been shown to not cleave DNA sequences with methylated cytosine bases, but sequences in and adjacent to transcribed regions in the DNA are generally not methylated. A certain endonuclease, *Pst*I, which is a methylation sensitive enzyme has been used to generate a genomic library of size-fractionated recombinant DNA that is 80 - 85% single copy DNA sequence. This map contains nearly 450 qualitative trait markers therefore speeding up the introduction of these characters into a cultivar or from one cultivar to another given that the marker and gene are closely linked.

The investigations identifying quantitative traits in the soybean genome have been limited. Seven markers for the quantitative trait of hard seededness have been found and these genomic regions explain a high percentage in the variation observed for this trait which affects properties such as germination rate , viability, and quality of stored seed (Shoemaker 1994). One locus associated with conditioning seed coat color has been shown to account for 32% of the variation of hard seededness. Additional QTLs affecting agronomically important traits such as maturity, seed fill, stem diameter, stem

length, canopy height, leaf width, and leaf length have been found using F_2 -derived lines and 150 markers. The variation explained by each marker was in the range of 16 - 24 %. Quantitatively inherited traits such as protein and oil content were evaluated with two hundred fifty-two markers in an F_2 derived line. Eight markers were significant for protein and nine for oil. Individual markers accounted for up to 38% of the variation in fatty acid composition, which is a direct measurement of soybean quality (Shoemaker 1994).

Allen (1994) remarked upon the usefulness of a molecular map and its significance in plant breeding. Using a molecular map, plant breeders can identify, manipulate, and complement traits to their maximum advantage. Indirect methods of trait selection such as morphological, biochemical, or molecular markers (such as RFLP's or simple sequence repeats) may be more desirable in many ways than direct trait selection through phenotype. Several reasons include early identification of individuals in the early growing stages for elimination or to recognize appealing individuals for crossing prior to flowering; better measurement of quantitative trait expression; and facilitating the selection of several traits simultaneously. Work with quantitative traits is slow and difficult, but can be improved with molecular mapping. In addition to indirect methods of selection, molecular markers can provide information about the number, effect and chromosomal location of genes affecting a particular trait, interaction among genes affecting a trait, and the function of genes in different environments.

Young (1996) describes quantitative trait loci mapping as an effective way to study traits that cannot be fitted to simple Mendelian ratios and are controlled by multiple loci. In multiple locus traits, environment and gene by gene interaction play an important

role in phenotypic expression, also properties such as additive and non-additive gene action are uncovered. DNA markers have made it possible to more effectively locate QTLs. DNA marker genotypes and phenotypes of interest are analyzed in a suitable population by splitting the individuals into classes according to genotype. Mean and variance parameters are calculated and evaluated among the phenotypic classes. A significant deviation among phenotypic classes where there are DNA marker genotype differences suggests that the DNA marker is related to the trait of interest.

In an experimental test done by Mansur *et al.*, (1993), linkages between RFLP markers and QTLs of agronomically important traits such as maturity, plant height, lodging, and seed yield were studied in 284 F₇ derived recombinant inbred lines (RIL's) acquired from the crossing of the soybean cultivars of Minsoy and Noir 1. Extreme phenotypes from each trait were bulked, digested with restriction enzymes, and Southern blotted using radioactive RFLP probes. When an RFLP marker was linked to a QTL, one parental allele predominated in the bulked DNA from a given phenotype, and the other parental allele was associated with the opposite phenotype. When equal amounts of radioactivity was associated with each allele in a given bulked lane, linkage was assumed to be absent. All four soybean traits were linked to marker A397, plant height was linked to R79, and all traits but lodging were shown to be linked to marker A60b. No evidence was shown for linkage of maturity, lodging, or seed yield to A60a, or of lodging to R79.

In another experiment by Lark *et al.*, (1994), the same population of 284 RIL from a cross of 'Minsoy' and 'Noir 1', were used to study quantitative trait values for protein and oil and for maturity. QTL for these traits were linked to several RFLP loci, including R183 for oil and protein content, and R79 for maturity. Noir 1 and Minsoy

differ in maturity by 2 days, whereas RIL plants with extreme phenotypes differed by 22 days. Values of seed oil for Minsoy and Noir 1 were 17.4% and 17.8% respectively, and for the extreme phenotypes, they were 15.5% and 19.6%. Similar differences in variation were shown for seed protein. Both late and early maturity appeared to be linked to R79. Low oil/high protein appeared to be linked to R183 but not to R79. High oil/low protein does not show significant linkage to either R79 or R183. Oil and protein were shown to be inversely related in this RIL population with a correlation of -0.934 . Earlier data suggested that different allele distributions were associated between trait values of maturity and for protein and oil. The distributions from the above study suggest that maturity effects are controlled by a locus acting independently and additively, but seed oil and protein are not independent.

In an investigation by Mian *et al.*, (1996), molecular markers associated with seed weight in two soybean populations were analyzed. Seed weight is an important component in soybean seed yield as well as an important trait in food-type soybeans. One hundred twenty F_4 -derived lines of 'Young' x PI416937 (Pop 1) and one hundred eleven F_2 -derived lines of PI97100 x 'Coker 237' were mapped with RFLP markers to identify QTLs conditioning seed weight across 3 locations and 2 years. the objectives of this study were to detect QTLs conditioning seed weight in soybean and to investigate the consistency of QTLs across populations and environments. A total of 155 RFLP markers were used to construct the genetic map of Pop1. One hundred forty markers were genetically linked covering a map distance of 970 cM. In Pop 2, there were 143 linked markers covering 1600 cM. In Pop 1, the parents did not differ in seed weight while the progeny differed by as much as 93 mg/seed. The parents of Pop 2 differed by

19 mg/seed and the progeny differed by 44 mg/seed. Seven independent marker loci were associated with seed weight in the combined analysis and six were identified in all three environments. Thus the level of consistency of the QTLs for seed weight across environments was very high. Individually the seven marker loci explained between 5 and 22% of the variation in seed weight. If combined, the seven marker loci would explain as much as 73% of the total variation in seed weight. The heritability of seed weight in this population was 90%. In Pop 2, nine independent loci were identified by single factor ANOVA as being associated with seed weight in the combined analysis. Six of these marker loci were consistent across all three environments and 2 were detected in 2 of the 3 environments. Of the eight marker loci detected in 1994, seven were also detected in 1995, showing the consistency level of significant QTLs across years to be very high. Individually the loci explained between 5 and 11% of the variation in seed weight. The nine loci combined would explain 74% of variation in seed weight. The heritability in this population was 91%. Pop 1 showed a positive correlation with seed yield and a negative correlation with maturity. There appeared to be no association between plant height and lodging. Pop2 had a positive correlation with plant height, lodging, and seed yield, but no relationship with maturity. The inconsistencies in the relationships of seed weight with other traits, between these two populations may be due to the fact that Pop 2 was segregating for growth habit and Pop 1 was not. The high level of consistency of marker loci across locations and years indicates that an effective marker assisted selection program is feasible for seed weight in soybean.

A population of sixty F₂ plants from a cross between a *G. max* experimental line, A81-356022 and a *G. soja* Sieb. and Zucc. plant introduction, PI 468916 was used in an

experiment by Diers *et al.*, (1992). The objectives were to present an expanded soybean RFLP map and to identify quantitative trait loci in soybean for seed protein and oil content. The map contains 252 markers, 31 linkage groups, and 2147 cM. The average distance between two marker loci is 8.5 cM. There were eight markers significantly associated ($P < 0.01$) with protein. Individually the markers explained between 12 and 42% variation in protein. Nine markers were significantly associated ($P < 0.01$) with oil content. Between 18 and 43% of variation in oil content was accounted for by individual markers. Alleles significantly associated with increased levels of one trait, protein or oil, were usually correspondent with lower levels in the other trait. This is consistent with the negative correlation generally found between these traits in soybean.

In research conducted by Lee *et al.*, (1996), a population derived from 'Young' and PI 416937 was evaluated with RFLP markers to identify QTL related to plant height, lodging, and maturity. Field data was collected over four locations. The genetic map consisted of 137 RFLP loci, 31 linkage groups, and more than 1600 cM. The parents differed by 45 cm in plant height where in the progeny, plant height differences were 71-127 cm. Fourteen RFLP markers associated with plant height were found in the combined analysis over all four locations. At all loci, except B1t043, the Young allele increased plant height. Young averaged 109 cm in plant height compared to 64 cm for PI 416937. The amount of variation explained by eleven independent markers sums to 68%. Heritability for plant height was 90.8 %. Young and PI 416937 differed by a lodging score of 0.5; however, the lodging scores among the progenies varied from 2.1 to 4.4. Fifteen RFLP loci were associated with lodging over all four locations. Eight of these were genetically independent and at six of the eight loci, the PI 416937 allele was

associated with increased plant lodging. The total amount of variation explained by the eight independent markers was 61%. Heritability of lodging was 82.3%. The parents differed by three days in maturity and transgressive variation of 36 days occurred among the progeny. Eight markers were detected over three locations and five of these eight were independent loci. At all five loci, except Blt043, the Young allele was associated with late maturity. Summed together, these five independent markers explained 48.8% of the total variation for maturity. The heritability for maturity was 93.1%. Increased plant height showed a positive association with increased lodging and later maturity. A positive correlation was also found for late maturity and increased lodging. At three RFLP loci the allele associated with later maturity was associated with increased height.

Efforts to identify QTL for seed quality traits have also been initiated. Lowering the content of linolenic acid, a fatty acid whose oxidation gives off a bad odor and taste, is a major goal in some plant breeding programs. In an experiment by Brummer *et al.*, (1995) the *fan* locus was linked with two RFLP markers. An F₂ segregating population was developed by crossing a low linolenic individual of *Glycine max*, C1640, to a high linolenic content soybean of *Glycine soja*, P1 479.750. The *fan* locus controlling this trait is thought to be inherited as a QTL because there were no clearly defined phenotypic classes. DNA from the F_{2:3} seeds was digested with five restriction enzymes: *Dra* 1, *Eco* R1, *Eco* RV, *Hin* d111, and *Taq* 1. Gel electrophoresis was run and genomic clones from the public soybean map were used for Southern blotting. A linkage map was developed using MapMaker v.3.0 with a logarithm of the odds ratio (LOD) at 3.0 and a recombination frequency of 50.0%. A USDA/ISU map was compared to a map with markers that identified the same size restriction fragments in the C1640 X P1 479.750

population and was used as a framework to align the two maps. MapMaker-QTL was used to more precisely place the *fan* locus on the map by assessing the likelihood at every position and then placing the locus on the linkage map at the position with the highest LOD score. Previous research has shown that the *fan* locus is the only locus controlling differences in linolenic acid content, but the progeny in this test did not fit the 1:2:1 ratio expected for a qualitative trait. Also, *G. soja* being one of the parents and of different genetic background, may have more than one locus responsible for linolenic acid content (Pantalone *et al.*, 1997). Regression analysis implied significant association of four RFLP markers with linolenic acid content: pB153, pB124, pB194-1, and pA183, and MapMaker-QTL determined that *fan* is most likely linked between markers pB194 and pB124

2.2 SIMPLE SEQUENCE REPEAT (SSR) MARKERS

As mentioned by Cregan *et al.*, (1994), the original idea of a DNA sequence repeat as a marker was a locus of Variable Number Tandem Repeats or VNTR. These sequence repeats or microsatellites were core sequence repeats varying in length from 11 to 60 base pairs flanked by endonuclease restriction sites. PCR was used to amplify these fragments which were then separated by gel electrophoresis into different fragment lengths. It was then suggested that maybe repetitive DNA sequences of only 2 - 5 core bp repeats would reveal higher levels of length polymorphisms. These shorter repetitive sequences are termed simple sequence repeats or SSRs and have been found to be abundant in many genomes. They were first extensively used in the human genome

linkage map, in which 339 of the markers are microsatellites and constitute more than 50% of the markers. SSR markers are placed on genetic maps in relation to other RFLP, DAF, RAPD and morphological markers.

In an article by Akkaya *et al.*, (1993), SSRs were also referred to as sequence tagged microsatellite sites or STMS. The function and the mechanism by which SSRs are created is unknown, but the number of repeats, “n”, varies from individual to individual giving easily distinguishable length polymorphisms. Their findings from analyzing seven loci revealed six to seventeen different size alleles in forty-three diverse soybean lines which is greater than that found in humans. An investigation of ten loci exposed four to eleven different size alleles with an average of 6.8 bp. Repeated sequences in soybean range in length from twenty-four to ninety-eight bp with an average of approximately fifty-seven bp in the seven loci studied. Of these sequences, AT appears to be the most common core repeat and the most polymorphic.

In a report done by Rongwen *et al.*, (1995), microsatellite markers were evaluated in their ability to determine the uniqueness of new soybean cultivars. The DNA sequences flanking microsatellites are generally conserved among individuals of the same species, allowing for selection of primers for PCR to be chosen, for the amplification product. The large variation in the number of tandem repeats produces highly polymorphic fragment lengths from genotype to genotype which are detected by PCR. Ninety-six cultivars were used to test the number of alleles found at three (AT)_n/(TA)_n loci and four (ATT)_n/(TAA)_n loci. Between 11 and 26 alleles were found at each of the seven loci. Lee and Pickett showed very similar SSR allelic profiles, but are very closely related being 93.4% identical which therefore accounts for their allelic similarity. The

incidence of “stutter” bands when using di-nucleotide repeats did make it difficult to determine the correct nucleotide length and so it was concluded that tri- and tetra-nucleotide sequences are preferable for fingerprinting.

Akkaya *et al.*, (1995) analyzed a soybean linkage map with SSR markers that had previously been mapped with RAPD and RFLP markers. The mapping population was an F₂ population of a cross between Clark and Harosoy. The purpose was to determine the map locations of SSR markers in relation to 13 loci controlling pigmentation and morphological traits, 7 isozyme loci, and a total of 118 RFLP and RAPD markers, and to ascertain if microsatellites were randomly distributed throughout the soybean genome. Forty microsatellite sequences available through GenBank, that were at least 20 bp long, were used. Thirty-four of the forty SSR loci were associated with 18 different linkage groups, while the remaining six were unlinked. Microsatellites were shown to be linked to five of the seven pigmentation traits, four of the six morphological traits, and three of the seven isozymes. Linkage group 13 was shown to contain nine markers, six of which were SSRs. The original map was increased by 40% with the addition of the 40 SSR markers. Overall, microsatellite loci appeared to be scattered randomly across the genome, but a significant deviation from the expectation, based on a Poisson distribution, was seen. However, 85% of this deviation was the result of the cluster of SSRs in linkage group 13, and so it was concluded that SSRs are spread in a random fashion throughout the soybean genome.

In an experiment outlined by Morgante *et al.*, (1994), sixty-one cultivars of soybean were used to map seven SSR markers. The cultivars included fourteen *G. soja* lines, twenty-six *G. max* cultivars, and twenty-one Plant Introductions mostly belonging

to the northern germplasm group. All seven markers showed at least four alleles at each locus, the average being 7.75, which may be an underestimate due to low resolution of two bp differences on agarose gels. The effective number of alleles, which is a measure of diversity that takes into account the number and frequency of each allele, was 2.57 and the level of differentiation showing the proportion of the population that would carry different alleles at a locus, was 0.62. Both measures of allele number and differentiation showed that allelic variation decreased with increasing levels of domestication while the wild soybean *G. soja* showed allelic variation as high as 90% at one locus. Of particular interest, is the fact that six of the seven markers used were (AT)_n repeats, because based on sequences based on databases, (AT)_n appears to be the most frequent type of repeat in plants. According to Akkaya *et al.*, (1992), there is a significant lack of (CA/GT)_n repeats in plants when compared to mammals, therefore (CA/GT)_n repeats will not give sufficient polymorphisms. A subset of 19 of the 61 lines used by Morgante *et al.*, (1994) showed a higher variation for SSRs than RFLP markers with the average number of alleles decreasing from 4.25 with SSR markers to 2.15 with RFLP markers.

In a study outlined by Yu *et al.*, (1993), the soybean mosaic virus (SMV) gene, *Rsv*, and the gene for anthocyanin pigmentation, *wl*, were mapped using RFLP's, SSRs, and morphological markers. SMV is one of the most common diseases in soybean production in the world resulting in seed quality reduction and serious yield loss, which makes the mapping of the SMV resistance gene an invaluable tool for breeders. A resistant line, P1 96983, was crossed with a susceptible line, Lee, and the F₂ generation mapped. Genes for pigmentation and SMV were shown to segregate with a 1:2:1 ratio indicating that they are both inherited in Mendelian fashion. *Rsv* was linked to the SSR

marker SM176, and to the RFLP markers pA186 and pK644a at 0.5, 1.5, and 2.1 cM respectively. The *wl* gene was also linked to this group at 48.8 cM distance from pK390.

Weising *et al.*, (1998) recounts the occurrence and frequency of microsatellites in the plant genome as differing considerably from that of animals. While the [CA]_n repeat is frequent in animals, it is comparatively rare in plants. In plants, stretches of [AT]_n, [GA]_n, [GAA]_n, and [TAA]_n repeats are the most common di- and tri- nucleotide sequences, respectively. There has been no correlation reported between microsatellite size and genome size. SSRs are relatively rare in chloroplast and mitochondrial genomes.

There are many ways in which microsatellites have been used as molecular markers, a few of them are listed here: microsatellite- primed PCR with unanchored primers, microsatellite primed PCR with anchored primers, and locus-specific microsatellite analysis using flanking primers. With the use of microsatellite -primed PCR with unanchored primers, microsatellite complementary oligonucleotides were used as the primers and if inversely repeated SSRs were within a reasonable distance, they were also amplified. The authors found that tetra nucleotide repeats were most efficient and that banding patterns were of higher complexity when separated on polyacrylamide gels, even small sized (<300 bp) were well resolved under these conditions. Vogel *et al.*, (1997), remarked that unanchored AT rich tri-nucleotides were not distinct on gel separation probably due to poor primer anchoring to the numerous target sites in the genome template. Also unanchored primers frequently anneal nonspecifically and length variation is not clear making this unsuitable for identifying differences within a species. In microsatellite-directed PCR using anchored primers, a 3' or 5' sequence of nonrepeat

bases was anchored to the SSR marker. It was predicted that anchored di-nucleotide based primers would generate distinct bands from individual loci. Weising *et al.*, (1998), showed that anchored GA and GT sequences produced the most useful banding profiles. There are obvious advantages of anchored over unanchored, some of which are, anchoring assures annealing of the primer only to the ends of the microsatellite, thereby reducing internal priming and smears on the gels; and by using a specific anchoring sequence, a specific loci can be targeted reducing the high number of PCR products. PCR amplified individual microsatellites using flanking primers (STMS) produce codominant markers that are specific for one locus. Even fragment size differences of <4 bp can be detected with this method using polyacrylamide gels; in many situations though, size differences are generally large enough to be resolved on high percentage agarose gels with ethidium bromide staining. The high number of polymorphic loci and the numerous dispersal of microsatellites throughout the genome makes SSR based primer systems, particularly STMS, an extremely useful tool in genetic mapping and genotype identification on the soybean.

Cregan *et al.*, (1999) reported that a number of RFLP based genetic maps of soybean have been developed over the past 10 years. However, since the soybean is an ancient polyploid, RFLP probes usually hybridize and map to more than one loci in the genome. The polymorphic potential of an RFLP probe is primarily a function of the frequency of the two alleles at each locus to which the probe maps to. SSR markers, however, are single locus markers with multiple alleles. The possibility of polymorphisms detected with SSR markers is dependent on the number of alleles and their frequency. The objectives of the work reported by Cregan *et al.*, (1999) was to

develop and map a large set of SSR markers. A total of 606 SSR loci were mapped in one or more of three populations: the USDA/Iowa State *G. max* x *G. soja* F₂ population, the University of Utah Minsoy x Noir 1 recombinant inbred population, and the University of Nebraska Clark x Harosoy F₂ population. Many SSR loci were segregating in two or all three populations. Therefore, the group was able to align the 20+ linkage groups derived from each of the three populations into a consensus set of 20 homologous linkage groups presumed to correspond to the 20 pairs of soybean chromosomes. In summary, microsatellite markers exhibit most of the properties desired in a marker system (Weising *et al.*, 1998):

- microsatellite markers are highly polymorphic
- microsatellite markers are codominant
- PCR generated microsatellites allow unambiguous separation of alleles, allowing for a precise calculation of allele frequencies
- microsatellites are abundant and evenly dispersed throughout the genome
- microsatellites are probably selectively neutral
- microsatellites are fast and easy to detect, using only nanograms of DNA
- the results are highly reproducible
- information is easily exchangeable between laboratories

2.3 HERITABILITY AND GENETIC VARIANCE

Heritability of a trait is determined as the ratio of genotypic variance, σ_g^2 , to the phenotypic variance, σ_{ph}^2 . Genotypic variance is the variation caused by genetic differences among individuals. The sum of additive, σ_A^2 , dominance, σ_D^2 , and epistatic, σ_I^2 , variances give genotypic variance. Phenotypic variance can be defined as visible differences among individuals and can be expressed as the sum of environmental variance or experimental error, σ_e^2 , genotype x environment interaction, σ_{ge}^2 , and genotypic variance. Genotype x environment interaction is the failure of genotypes to perform the same relative to each other when they are evaluated in different environments and can be expressed as the sum of genotype x location (σ_{gl}^2), genotype x year (σ_{gy}^2), and genotype x location x year (σ_{gly}^2). Heritability can be classified as either broad sense or narrow sense. Broad sense heritability is the ratio of total genotypic variance, including additive, dominance, and epistatic, to phenotypic variance, σ_g^2/σ_{ph}^2 . Narrow sense heritability is the ratio of additive variance to phenotypic variance, σ_A^2/σ_{ph}^2 (Fehr 1987).

In an investigation by Pantalone *et al.*, (1996), heritability of soybean fibrous roots and genetic correlation with other agronomically important traits was assessed. PI 416937 has been shown to exhibit higher nitrogen fixation, which could be due to a high phenotypic fibrous root score positively correlated with other root traits such as nodule number and nodule dry weight. A cross between PI 416937 (fibrous root) and Lee (non-fibrous root) was made and $F_{2:3}$ and $F_{2:4}$ progeny were evaluated for two years over two

environments to determine heritability estimates. The root score heritability estimate of 0.39 was low, similar to that for yield (0.36). A significant ($P=0.10$) positive correlation ($r^2=0.42$) was found between root score and seed protein, signifying some association between the fibrous root trait with seed protein concentration.

In an article by Cober *et al.*, (1997), the heritability of soybean seed shape and size was evaluated using parent/offspring regression, parent/offspring correlation, and realized heritability. The correlation between seed size and seed shape was also determined. For cultivars used for food, round seeds are desirable and understanding that heritability would be beneficial to the development of new cultivars. Three crosses between X3834, X3852, X3854 and parents, 9042 and 9092 were made. In the F_2 and $F_{2:3}$ generations evaluated for seed size, the parent/offspring regression coefficients were 0.19, and 0.51; the parent/offspring correlation coefficients were 0.29 and 0.56; and the realized heritability estimates were 0.26 and 0.50. From this data, heritability was concluded to be low for seed size. Estimates for heritability for seed shape were 0.55 to 0.75 for parent/offspring regression coefficients; 0.49 to 0.76 for parent/offspring correlation coefficients; and 0.59 to 0.79 for realized heritability. These statistics show seed shape to be a moderately heritable trait indicating that parent seed shape is a good predictor of progeny seed shape. Correlations from seed shape and size were not significantly different from zero.

3. INTRODUCTION

3.1 MOLECULAR MARKER STUDY

In plant breeding and genetics, markers include any phenotype detected as a distinct protein or segment of DNA or RNA. It can be an unrecognized sequence with no function, or it can correspond to a genetically expressed sequence. It may just function as a reference point on the genome. Although a molecular marker is physically a DNA molecule, this does not mean that it operates and can be used as a genetic marker. Some DNA sequences are altered during development by methylation, causing changes in expression or sensitivity to their mechanism of detection (Grattapaglia 1994).

The introduction of recombinant DNA technology and the discovery of the Polymerase Chain Reaction (PCR) have caused a shift from genetic analysis based on the presumption of genotype from phenotype, to direct genetic analysis based on variation at the molecular DNA level. Prior to the 1960's, markers used in plant genetics and breeding were the actual phenotype of the plant such as dwarfism, flower, seed or leaf color, and morphology. These morphological markers were the basis of the theoretical development of genetic segregation and linkage and are what established the first versions of genetic maps. However, due to their limited number and the fact that morphological markers were rarely close to any economically important genes, they were of little practical use except in a few model plant species such as pea, maize, and tomato (Grattapaglia 1994).

With the advent of modern techniques in molecular biology, several methods for the detection of DNA polymorphisms have been discovered. The discovery and use of restriction enzymes and DNA hybridization allowed the analysis of Restriction Fragment Length Polymorphisms (RFLP's). Along with the discovery of PCR, by which fragments of DNA are amplified using a thermostable DNA Polymerase, several classes of molecular markers based on this technology have been identified, such as Random Amplified Polymorphic DNA (RAPD), DNA-amplification fingerprint (DAF), and microsatellites also known as simple sequence repeats (SSRs) (Grattapaglia 1994).

Some advantages of molecular markers as genetic markers are:

- A large number of genetic marker loci can be followed in segregating populations allowing the construction of a genetic map. The polymorphism information content of such marker loci is generally high while morphological markers typically have a low polymorphism information content.
- Molecular markers are generally neutral in relation to phenotypic effects with minimal or no epistatic or pleiotropic effects while morphological markers frequently restrict normal plant development, such as dwarfs, albinos, etc., and show epistatic effects that can complicate the correct identification of genotypes.
- Molecular markers are generally co-dominant, containing more information than morphological markers, which generally show dominant or recessive inheritance (Grattapaglia 1994).

Soybean, being an autogamous or self- pollinating crop, is essentially homozygous at many loci and therefore more difficult to map because of lack of

polymorphism, consequently, SSRs were chosen as markers because of their highly polymorphic nature (Grattapaglia (1994).

3.2 FIELD BREEDING STUDY

Soybean, *Glycine max* (L.) Merr, is in the Fabaceae or the Leguminosae family with a *Pisum* type flower. Soybean is a coarse, bushy annual plant with stems up to 1.8 meters sometimes growing vine-like. Stems and leaves are pubescent, either gray or tawny, and leaves are pinnately trifoliate, exhibiting a broadly ovate to elliptic-lanceolate shape. Soybean is a subtropical plant and cannot withstand excessive heat or severe winters tolerating a temperature range of 5.9° to 27° C. It is a short day plant and grows best on fertile, well-drained soils, tolerating a wide range of soil conditions although it prefers a pH of 6.0 – 6.5. Soybean can endure an annual precipitation of 3.1 to 41.0 dm and a pH of 4.3 to 8.4. Soybean should not be fertilized with nitrogen unless the soil lacks the proper nitrogen-fixing bacteria needed for nodulation (www.hort.purdue.edu).

Soybean is propagated by seed and responds well to planting with thorough cultivation in fall or early spring to provide a deep loose seedbed. Soybean also responds well when properly sown in no-tillage systems. Seeds are usually planted at a depth of 2.5 to 5 cm, and spaced about 2.5 cm apart. Close spacing encourages rapid growth and aids in weed control, but spacing too close can seriously increase lodging, which makes combining difficult and reduces yields. The seed is ready for harvest when it matures. Seed on a soybean plant primarily matures at the same time. Dropping of the leaves and drying of the stems accompany seed maturity. Combining can cause yield losses of 10 to

20 % due to seed splitting and mechanical damage. Damage may increase as seed moisture drops below 12% (www.hort.purdue.edu).

The highest priority in soybean breeding is increasing seed yield. Seed yield is a quantitative trait controlled by many genes and strongly influenced by the environment. Heritability of seed yield is the lowest and most variable of the major agronomic traits taken into account during cultivar development, with estimates ranging from 3 to 58% (Brim *et al.*, 1973). Over the years, considerable genetic improvement for seed yield has been attained, with an average rate of yield increase of 15.1 kg/ha per year, or 0.6% per year since the 1920's. Every year, new public and private cultivars are released that are higher in seed yield than existing cultivars (Fehr 1987).

Other major characters that are being developed in breeding programs are maturity, lodging resistance, pest resistance, and seed quality, composition, and size. Recently, marketing the protein and oil content of the seed has become important, depending on the intended use of the seed (Fehr 1987). Protein and oil are quantitative traits with a heritability of about 65% (Brim *et al.*, 1973). Seed size, heritability about 70% (Brim *et al.*, 1973), is an important trait depending on use. Large seeds are preferred for use as a vegetable and for tofu, while smaller seeds are preferred for some foods such as a Japanese fermented food called natto (Fehr 1987).

The population in this experiment was developed from a cross between Williams and Essex. Essex is a southern cultivar, originating as an F₇ line at the Virginia Agricultural Experiment Station from the cross 'Lee' x S5-7075. S5-7075 is a selection from the cross, N48-1248 x 'Perry'. N48-1248 is a selection from the cross 'Roanoke' x a selection from 'Ogden' x 'CNS' (Smith *et al.*, 1973).

Characteristics of Essex include high seed yield, excellent standing ability, and good seed quality. Seeds have a buff hila, yellow cotyledons, a yellow seed coat (Smith *et al.*, 1973), and a one hundred seed weight of 11.1 g (www.ars-grin.gov). Essex seeds are free from mottling and have a good resistance to purple seed stain. Essex is resistant to bacterial pustule, several races of downy mildew, and frogeye leafspot and moderately resistant to phytophthora rot (Smith *et al.*, 1973).

Williams is a northern variety developed in a cooperative breeding program of the U.S. Regional Soybean Laboratory and the Illinois Agricultural Experiment Station. Williams is the progeny from an F₅ plant selected from the cross of 'Wayne' x L57-0034. The F₂ through F₅ generations were grown in southern Illinois under conditions of severe stress on seed quality and pod set. Selections were made each year for good seed quality, pod set, plant vigor, and lodging resistance. The parent line L57-0034 was a selection from the cross 'Clark' x 'Adams' (Bernard *et al.*, 1972).

Pods of Williams have a tan pod wall and seeds have a shiny yellow seed coat and black hila (Bernard *et al* 1972), and a one hundred seed weight of 19.0 g (www.ars-grin.gov). Williams average slightly higher in yield over the region tested and is adapted to the southern part of the North Central States. Williams has a moderate susceptibility to phytophthora rot, high resistance to bacterial pustule, high susceptibility to downy mildew, and a moderate susceptibility to iron chlorosis in soils with a high pH (Bernard *et al.*, 1972).

The F₂ seeds of our Essex x Williams population were sent to Costa Rica for three generations of advancement, via single seed descent, from July 1998 to April 1999, and the F_{4:6} seeds were grown in Knoxville and Springfield, TN in 1999. Quantitative and

qualitative traits were scored on the $F_{4:6}$ lines and genetic variance, and heritability among traits was computed.

4. MATERIALS AND METHODS

A cross between the northern soybean cultivar Williams and the southern cultivar Essex was made by Ali Ustun (1999). Several traits were recorded from the F₂ population and using SSRs, we attempted to link these traits to 10 SSR markers that were found to be polymorphic.

To initially identify polymorphism for potential QTL detection, each SSR primer pair was tested on DNA of the Essex and Williams parents and six random progeny. Those primer pairs that produced different sized PCR products in the parents and which segregated in the progeny were then screened among all of the progeny and scored by running each product on polyacrylamide gels.

4.1 DNA ISOLATION AND CHARACTERIZATION

DNA from new leaves of soybean was collected from an F₂ generation of a cross between Williams and Essex. The leaves were stored in -20° C until used for DNA extraction.

The protocol for DNA extraction and isolation from leaves of F₂ plants is outlined below from Dellaporta (1983):

1. Leaves were ground with liquid nitrogen and insoluble PVP (polyvinylpolypyrrolidone).

2. Cold, ground leaf tissue was transferred into a 50 ml plastic centrifuge tube along with 15 ml of extraction buffer: 100mM Tris HCl, 50mM EDTA, and 500mM NaCl (pH 8.0) and 1 ml of 20% SDS, mixed well with vortex mixer
3. Mixture was incubated for 30 min to 1 hour in a 65 °C water bath.
4. 5ml 5M K Acetate (pH 8.0) was added, then mixed.
5. Mixture was packed in ice for 15-30 minutes as a cleaning step where SDS and K-Acetate form complexes and precipitated out.
6. The solution was poured through filter paper into clean plastic centrifuge tubes.
7. 10 ml of cold (-20° C) 99% - 100% isopropanol was added and mixed well.
8. Tubes were stored in -20° for 30 min. (overnight storage is possible)
9. Samples were centrifuged 10-15 min using the JA20 rotor @ 4° C and 12000 rpm.
10. The supernatant was poured out and the isopropanol was allowed to evaporate.
11. 1ml of chromosomal DNA buffer: 50mM Tris HCl and 10mM EDTA. (PH 8.0) was added. It is important that the buffer covered the pellet in order to dissolve the DNA.
12. Samples were put on ice or stored at 4° C overnight to dissolve the pellet.
13. Solution was pipetted into 1.5 ml eppendorf tube and centrifuged with eppendorf centrifuge for 10 min @ 12000 rpm to precipitate out polysaccharides and SDS complexes.
14. Solution was pipetted into new 1.5 ml eppendorf tube and 1/10 volume (100 µl) of 3 M sodium acetate (pH 7.50) added.
15. 500 µl propanol was added and gently but thoroughly mixed by inverting the tube.
16. Samples were centrifuged full speed for 2 min and supernatant was poured off.

17. Samples were washed once or twice with 70-80 % ETOH (750 μ l) to remove salts.
18. Samples were allowed to stand at room temperature for 5 min and then centrifuged full speed for 2 min.
19. ETOH was poured off and let evaporate.
20. Pellet was dissolved in 200 μ l of autoclaved distilled H₂O (or TE buffer) and stored at 4° C until use.

4.2 DNA QUANTIFICATION

Each extracted DNA sample was quantified using a fluorometer. Calf thymus DNA of a known concentration of 1000 ng/ μ l was diluted to 500 ng/ μ l and then used to calibrate the fluorometer. Two microliters of each sample was added to a cuvette containing 2 ml of 1X TNE (100mM Tris, 10mM EDTA, 2M NaCl) and Hoescht dye. Each sample was then accordingly diluted to 15 ng/ μ l.

4.3 POLYMERASE CHAIN REACTION (PCR)

A PCR master mix was made for each primer before amplification. The master mix contained, for each sample: 5.2 μ l dd H₂O, 1 μ l of 10X PCR Gold Buffer, 1 μ l of 200 μ M nucleotide mixture, 1.2 μ l of 25 μ M MgCl₂, 0.3 μ l of each the 20 μ M 3' and the 5'

SSR primer, and 0.08 ul of Amplitaq Gold DNA Polymerase. Then 8.75 ul of the master mix was added to each reaction tube along with 1.25 ul of the 15 ng/ μ l template DNA.

Cycling consisted of 1 cycle of 96° C for two minutes to activate the enzyme, 30 cycles each of denaturation at 96° C for 45 sec, 48° C for 45 sec for primer annealing, and 72° C for 1 min for transcription, and then, 1 cycle of 72° C for 7 minutes. PCR products were separated on polyacrylamide gel with silver staining. \

4.4 POLYACRYLAMIDE GEL ELECTROPHORESIS

Ten milliliter mini-gels were used to detect polymorphisms of at least 10 base pairs. The largest base pair difference was 28 base pairs, detected by Satt373. Forty milliliter gels were utilized to detect polymorphisms as small as 3 base pairs.

The 40 ml gels were run on the BioRad Protean II. Gels were formed by the following procedure: 40 ml of 5% polyacrylamide was added to a beaker; ammonium persulfate (APS) was added to the polyacrylamide at 150 ul/ 10ml 5% gel, and amount of Temed added was 10% of APS; for 40 ml of gel, 600 ul of APS and 60 ul of Temed was added at the same time while the gel solution was on a stirrer. The solution was then pulled into a syringe and poured between two large glass plates (each bottom plate has been treated with 5:1 ethanol and acetic acid solution and 10 ul of Bind Silane). Combs were added to each gel to form wells and then the gels were set to polymerize for 20 minutes. Once polymerized, the gels were put into the gel box and 1X TBE buffer was added making sure the top and bottom of both gels were in contact with the buffer. Three microliters of each PCR product and 3 ul bromophenol blue tracking dye was loaded into each well. A 25 bp DNA ladder was added to the first and last wells to serve as a

DNA fragment size standard. An electric current of 350 – 400 volts was then run through the gels for 3 or 4 hours, migrating each DNA sample according to length.

The 10 ml gels were run on a 10% mini gel (BioRad Mini Protean). A polyester film backing, with a hydrophilic and a hydrophobic side, was placed hydrophilic side up against the bottom plate of the mini gels. The gels were cast against the film. Nine or ten milliliters of a 10% gel solution with 140 μ l APS and 14 μ l Temed were used for mini gels. A syringe tip filter was best used when pouring 10% gels to filter out undissolved particles. The gels were then run at 180 volts for about 1 hour (Trigiano and Caetano-Anolles, 1998).

Once migrated all the way, the current was turned off and the gel apparatus was taken apart.

4.5 SILVER STAINING

The bottom glass plate with the gel bound to it was put in a tray with 10% acetic acid and put on a shaker for 10 minutes. Then dd H₂O was added two times for 5 minutes each to rinse the gel. A solution of silver nitrate (0.1g / 100ml dd H₂O) was poured over the gels, about 400 ml per gel, and allowed to shake for 25 to 30 minutes. After staining, the gels were rinsed again in dd H₂O for 10 to 30 seconds. A developing solution of sodium carbonate (3g / 100ml dd H₂O), which had been chilled to 4° C was then poured over the gels, about 400 ml per gel, and shaken until the bands appear (this usually took about 5 –8 minutes). Once the bands became dark enough, 10% acetic acid, which had been kept at 4° C, was poured over the gels to stop development. The gels were soaked in the acetic acid for 5 – 10 minutes, and then rinsed in dd

H₂O two times for 5 minutes each time. The gels were then allowed to dry prior to analysis (Bassam *et al.*, 1991).

Procedures were modified slightly for staining mini gels. The first rinsing was done three times for 2 minutes each time. For both mini-gels, only 150 ml of silver nitrate and sodium carbonate were needed. The very last step was soaking the gels in an anticracking solution for 5 to 10 minutes and then allowing to dry (Trigiano and Caetano-Anolles, 1998).

4.6 SCORING

The bands for each parent and progeny were scored according to the polymorphic bands. An upper band was assigned a score of 1, while a bottom band was assigned a score of 3. Any heterozygous progeny, showing both band lengths, were assigned a score of 2. Table A.3 is a list of all the microsatellite primers screened, and table A.4 shows the primer sequence for the ten primers that were polymorphic among the progeny. QTL statistical analysis of variance of scored bands was run via GLM (SAS) using a general model and a maturity covariate model.

4.7 FIELD BREEDING STUDY

In 1998 four F_{2:3} seeds from each of 208 individual F₂ plants were sent to Costa Rica over the winter to advance the population three generations to F₆. In July 1998, the four F_{2:3} seed were planted in a single hill for each packet. In September 1998, one F₃ plant was harvested from each hill and eight F_{3:4} seeds were collected from each plant. In October 1998, four F_{3:4} seeds from each packet were planted in a single hill. After

germination, the hills were thinned to one plant per hill. At harvest, all the seed from individual plants were threshed for $F_{4:5}$ seed. The seed were increased by growing a six-meter row of $F_{4:5}$ seed from each of the individual threshed plants. Each row was bulk harvested and the 208 bags of bulk harvested $F_{4:6}$ seed were shipped back to Knoxville TN.

In the field, 208 $F_{4:6}$ lines derived from F_2 plants from hybridization between Williams and Essex. The parental lines, and the checks used in the USDA uniform test program, 'Macon' (MG III), 'KS4694' (MG IV), 'Manokin' (MG IV-S), and 'Hutcheson' (MG V), were evaluated to identify or measure the phenotypes for flower color, pubescence color, maturity, yield, lodging, height, seed size, protein concentration, and oil concentration. There were two locations, Knoxville Experiment Station in Knoxville, TN and Highland Rim Experiment Station in Springfield, TN. Three replications were planted in one row plots, six meters in length. Genotypic variance was estimated by calculating variance components using Table A.5 outlined by McIntosh (1983). Phenotypic variance was calculated using the formula $\sigma_p^2 = \sigma^2/re + \sigma_{ge}^2/e + \sigma_g^2$. Heritability was then calculated with the formula $h^2 = \sigma_g^2/\sigma_p^2$. Genotypes were considered a random effect, representing a random sample of potential genotypes from the population. Environments were considered random, representing a sample of the available experimentation station field sites with unknown weather patterns for 1999. Statistical analysis were conducted using the MIXED procedure of SAS.

In early summer of 1999 plants were determined to have either white, purple, or segregating flower color. Beginning in September 1999, plants were measured for

maturity when 95% of the pods attained their final color. At this time, plots were recorded for pubescence color (tawny, gray, or segregating). Lodging was scored and assigned a score of 1 through 5. Plants with a 90° angle to the ground were assigned a score of 1. Plants that were 67.5° from the ground were assigned a score of 2, and plants with a 45° angle to the ground were assigned a 3. Plants with a 22.5° angle to the ground were assigned a 4 and plants with a score of 5 were lying completely prostrate. Half points were also assigned. Height of each plant was measured using a 5-foot PVC pipe marked in foot and inch increments and was converted to centimeters.

Individual rows were harvested seven days after their maturity date. Each row was then hand threshed and measured for yield and 100 seed weight. A 20 to 25 gram sample was then sent to National Center for Agricultural Utilization Research (NCAUR) Peoria, Illinois to measure oil and protein concentration of the seed via NIR.

5. RESULTS AND DISCUSSION

5.1 F₂ POPULATION QTL ANALYSIS

Measured traits from the F₂ and F_{4:6} populations were analyzed with the microsatellite marker scores from the DNA of F₂ progeny. Maturity, yield, seed protein concentration, seed oil concentration, seed size, and plant height were measured for both the F₂ and the F_{4:6} populations and lodging was measured only for the F_{4:6} populations. The F₂ mean and progeny range data are shown in Table 1. Maturity ranged from 115 to 173. Yield ranged from 6.7 to 201.6. Seed Protein ranged from 33.6 to 42.8. Seed Oil ranged from 22.1 to 24.6. Yield distribution for the F₂ population is shown in Figure 1 and seed protein distribution for the F₂ population is shown in Figure 2. Data were analyzed using two different models in SAS, a general additive-dominant model (soybean is an inbred crop with pure lines in which additive effects can be detected) and a covariate model with maturity as the covariate. Essex is a southern cultivar adapted to maturity group V growing regions such as Tennessee. Williams is a northern cultivar adapted to the mid-west growing region. Because of the different maturities of the parents, a wide distribution in maturity among progeny would be expected from this cross. Maturity is associated with the growth and development of a plant throughout the growing season. It is therefore possible that wide differences in maturities among the progeny may influence yield, height or other measured traits. Therefore in addition to the general model, which tested each trait individually without taking maturity into effect, we also used a maturity-covariate model to determine whether discovered QTLs still

Table 1. Mean and F₂ Progeny Range of Maturity, Yield, Seed Protein, and Seed Oil.

	Days to Maturity[†]	Yield	Seed Protein (% DW)	Seed Oil (%DW)
Mean	140.9	66.3	39.1	23.2
Progeny Range	115-173	6.7-201.6	33.6-42.8	22.1-24.6

[†] Days from planting to R8.

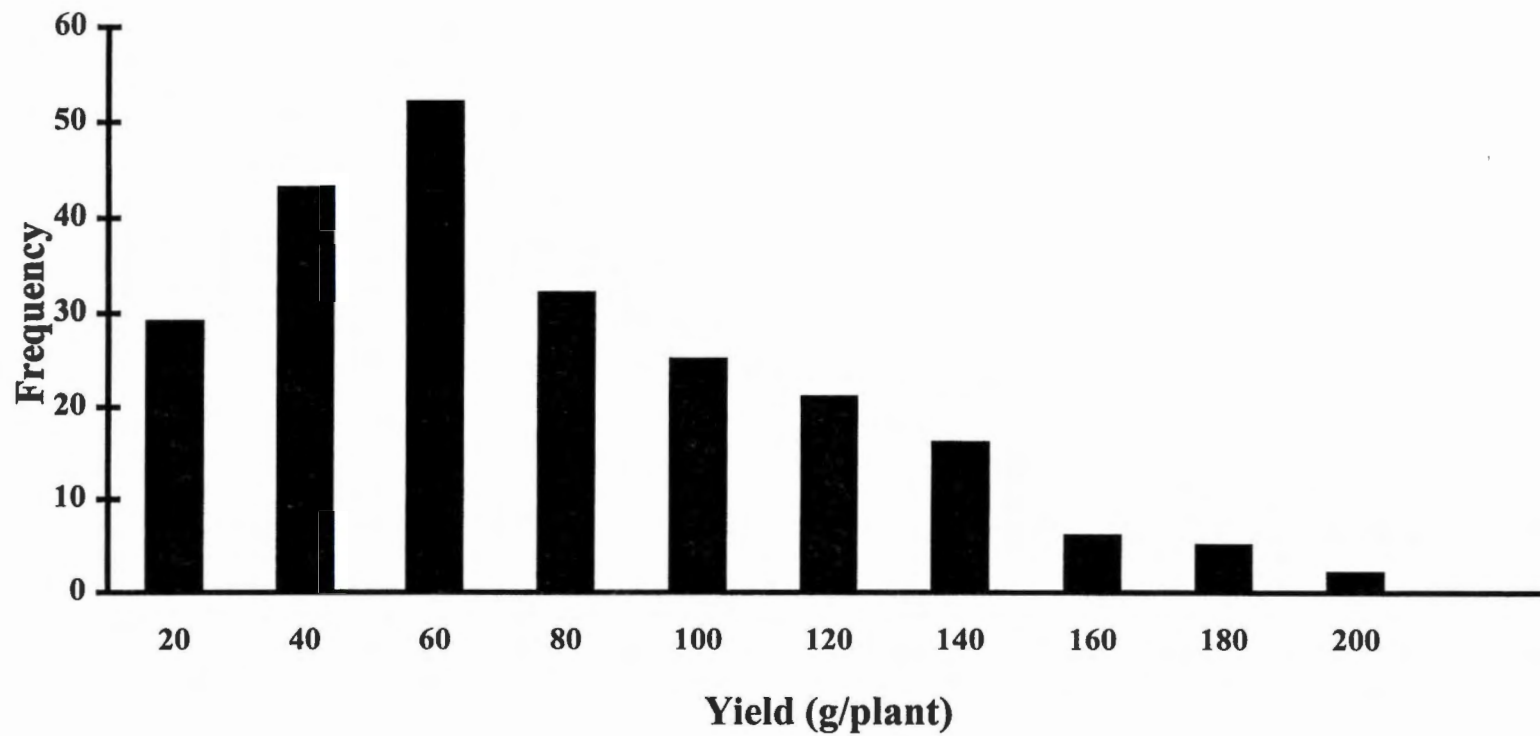


Figure 1. Frequency Distribution for Plant Yield of Essex x Williams F₂ Population

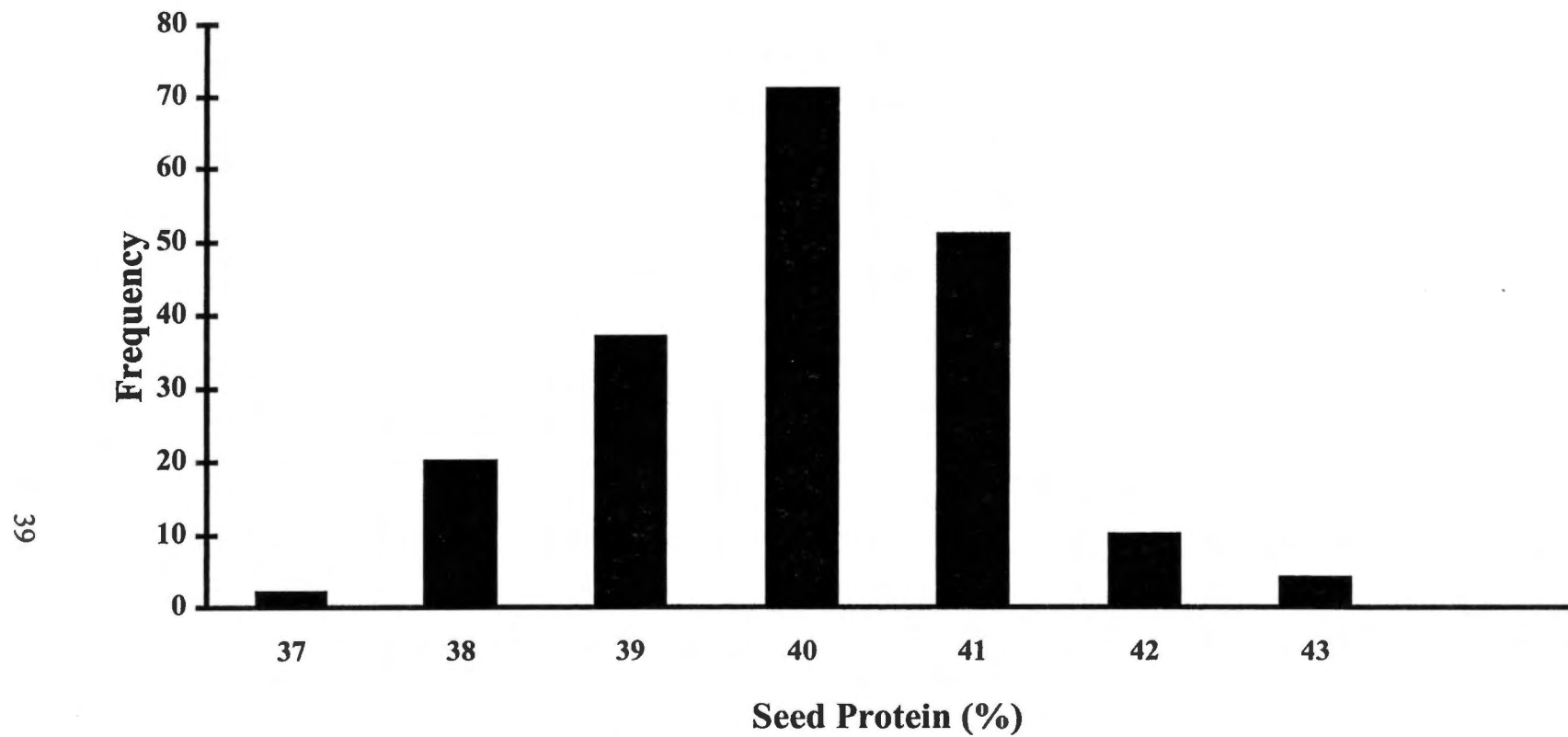


Figure 2. Frequency Distribution for Seed Protein Concentration of Essex x Williams F₂ Population

explained a significant portion of variation for each trait, after factoring out maturity and adjusting for its effects. The analysis of maturity covariance was conducted to improve the precision of the experiment by statistical control of variation among experimental units. By factoring out maturity in the covariate model, we removed the variation from experimental error and adjusted the treatment means for differences attributable to maturity obtaining a tighter model with a smaller error term (Rawlings 1988).

The general and the covariate model both estimate values for additive and dominant effects. The additive effects are measured by the additive genotype, or breeding value. Since it is the average effects of the parent's genes that determine the mean genotypic value of its progeny, the breeding value of an individual is judged by the mean value of its progeny and is equal to the sum of the average effects of the genes it carries. This summation is made over the pair of alleles at each locus and over all loci. Because the breeding value expresses the value transmitted from parents to offspring, it ensues that the expected breeding value of any individual is the average of the breeding values of its two parents. This value is also the expected phenotypic value (Falconer 1989). When a single locus is under consideration, the difference between the genotypic value and the breeding value is the dominance deviation. This dominance effect arises from the property of dominance among the alleles at a locus. Statistically, the dominance deviations are interactions between alleles, or within-locus interactions. They represent the effect of putting genes together in pairs to make genotypes (Falconer 1989). Both models also give values for the coefficient of determination, r^2 , for each independent variable analyzed separately.

The general QTL analysis model for maturity in the F₂ Knoxville population is shown in Table 2. QTLs are deemed to be significant for p-values ≤ 0.10 in this experiment. Satt540 is significant for maturity (p=0.0001) with significant additive effects (p=0.0001) and dominant effects (p=0.054). The r^2 value shows that marker Satt540 explains 10.6% of the variation in maturity among progeny. Satt52 is not significant for the marker or additive effects, but shows significant dominant effects (p=0.059) explaining 2% of the variation among progeny. Marker Satt50 is significant for maturity (p=0.0008) with significant additive effects (p=0.0002). The r^2 value shows that marker Satt50 explains 7.3% of the variation in maturity. Satt239, Satt373, Satt9, Satt353, Satt30, Satt251, and Satt14 were not significant for any of the values.

The QTL analysis of yield for the F₂ population is shown in Table 3. Marker Satt540 is significant (p=0.0022) for the general model and it shows additive (p=0.0047) and dominant (p=0.032) effects. Satt540 explains 6% of the yield variation in this model. The maturity covariate is significant (p=0.0001) for Satt540, nevertheless, the marker p-value is significant (p=0.086) showing that Satt540 is a true QTL for yield. Dominant effects were also significant (p=0.098) and the adjusted r^2 value shows that Satt540 explains 13.3% of the variation in yield in the maturity covariate model. Satt52 and Satt373 were not significant for the marker overall in the general model, but show some significant additive effects (p=0.070 for Satt52, p=0.067 for Satt373). The maturity covariate is significant for both markers (p=0.0001) in the adjusted model, but the marker p-values are not significant and there are no significant additive or dominant effects for either. Thus, significance in the general model was probably due to maturity effects and not to yield. Satt239 is significant (p=0.076) and also shows significant additive effects

Table 2. QTL Analysis of Essex x Williams F₂ Soybean Population for Plant Maturity at Knoxville, TN.

General model				
Marker	Marker P-value	r ² (%)	Additive P-value	Dominant P-value
Satt540	0.0001 ^{***}	10.6	0.0001 ^{***}	0.054 [†]
Satt52	NS	2.0	NS	0.059 [†]
Satt239	NS	NS	NS	NS
Satt373	NS	NS	NS	NS
Sctt9	NS	NS	NS	NS
Satt50	0.0008 ^{***}	7.3	0.0002 ^{***}	NS
Satt353	NS	NS	NS	NS
Satt30	NS	NS	NS	NS
Satt251	NS	NS	NS	NS
Satt14	NS	NS	NS	NS

†, *, **, *** Significant at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

NS, Not Significant at the 0.10 level.

Table 3. QTL Analysis of Essex x Williams F₂ Soybean Population for Plant Yield at Knoxville, TN.

Marker	General Model				Maturity Covariate Model				
	Marker P-value	r ² (%)	Additive P-value	Dominant P-value	Maturity Covariate	Marker P-value	r ² (%)	Additive P-value	Dominant P-value
Satt540	0.0022**	6.0	0.0047**	0.032*	0.0001***	0.086 [†]	13.3	NS	0.098 [†]
Satt52	NS	1.8	0.070 [†]	NS	0.0001***	NS	NS	NS	NS
Satt239	0.076 [†]	2.7	0.023*	NS	0.0001***	0.087 [†]	13.0	0.027*	NS
Satt373	NS	1.8	0.067 [†]	NS	0.0001***	NS	NS	NS	NS
Scrt9	NS	NS	NS	NS	0.0001***	NS	NS	NS	NS
Satt50	NS	NS	NS	NS	0.0001***	NS	NS	NS	NS
Satt353	NS	NS	NS	NS	0.0001***	NS	NS	NS	NS
Satt30	NS	NS	NS	NS	0.0001***	0.089 [†]	14.1	NS	0.052 [†]
Satt251	NS	NS	NS	NS	0.0001***	NS	NS	NS	NS
Satt14	NS	NS	NS	NS	0.0001***	NS	NS	NS	NS

[†], *, **, *** Significant at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

NS, Not Significant at the 0.10 level.

($p=0.023$) explaining 2.7% of the yield variation in this population. The maturity covariate is significant ($p=0.0001$) for Satt239 and so is the marker p -value ($p=0.087$) meaning that Satt239 is a true QTL for yield explaining 13% of the variation for yield. There are also significant additive effects ($p=0.027$). Satt30 is not significant for any p -values in the general model, but is significant for the maturity covariate. This tighter model was able to pick up significance for the marker ($p=0.089$) and for dominant effects ($p=0.052$) after factoring out maturity. Satt30 is therefore a true QTL explaining 14.1% of the variation in yield. None of the rest of the markers, Sctt9, Satt50, Satt353, Satt251, or Satt14 were significant for any p -values in the general or covariate models.

The QTL analysis of seed protein concentration for the F_2 population in Knoxville is displayed in Table 4. Marker Satt52 is significant ($p=0.064$) for seed protein and shows significant dominant effects ($p=0.021$). The maturity covariate is not significant. Satt52 could be a true QTL explaining 3% of the variation for seed protein concentration. Marker Satt251 is significant ($p=0.081$) and shows significant additive effects ($p=0.029$). The maturity covariate is not significant. Satt251 could be a true QTL, explaining 2.7% of the seed protein variation. The rest of the markers, Satt540, Satt239, Satt373, Sctt9, Satt50, Satt353, Satt30, and Satt14 were not significant.

The analysis of seed oil concentration for the F_2 population in Knoxville is shown in Table 5. Satt52 is significant ($p=0.062$) for seed oil and significant for dominant effects ($p=0.021$). The maturity covariate is not significant, therefore Satt52 could be a true QTL explaining 3% of the variation in seed oil concentration. Marker Satt373 does not have a significant marker p -value, but shows significant additive effects ($p=0.077$), which explained 1.8% of the variation. The maturity covariate is not significant. Marker

Table 4. QTL Analysis of Essex x Williams F₂ Soybean Population for Seed Protein Concentration at Knoxville, TN.

General Model					Maturity Covariate Model				
Marker	Marker P-value	r ² (%)	Additive P-value	Dominant P-value	Maturity Covariate	Marker P-value	r ² (%)	Additive P-value	Dominant P-value
Satt540	NS	NS	NS	NS	NS	----	----	----	----
Satt52	0.064 [†]	3.0	NS	0.021 [*]	NS	----	----	----	----
Satt239	NS	NS	NS	NS	NS	----	----	----	----
Satt373	NS	NS	NS	NS	NS	----	----	----	----
Sctt9	NS	NS	NS	NS	NS	----	----	----	----
Satt50	NS	NS	NS	NS	NS	----	----	----	----
Satt353	NS	NS	NS	NS	NS	----	----	----	----
Satt30	NS	NS	NS	NS	NS	----	----	----	----
Satt251	0.081 [†]	2.7	0.029 [*]	NS	NS	----	----	----	----
Satt14	NS	NS	NS	NS	NS	----	----	----	----

[†], *, **, *** Significant at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

NS, Not Significant at the 0.10 level.

Table 5. QTL Analysis of Essex x Williams F₂ Soybean Population for Seed Oil Concentration at Knoxville, TN.

General Model					Maturity Covariate Model				
Marker	Marker P-value	r ² (%)	Additive P-value	Dominant P-value	Maturity Covariate	Marker P-value	r ² (%)	Additive P-value	Dominant P-value
Satt540	NS	NS	NS	NS	NS	----	----	----	----
Satt52	0.062 [†]	3.0	NS	0.021 [*]	NS	----	----	----	----
Satt239	NS	NS	NS	NS	NS	----	----	----	----
Satt373	NS	1.8	0.077 [†]	NS	NS	----	----	----	----
Satt9	NS	NS	NS	NS	NS	----	----	----	----
Satt50	NS	NS	NS	NS	NS	----	----	----	----
Satt353	NS	NS	NS	NS	NS	----	----	----	----
Satt30	NS	NS	NS	NS	NS	----	----	----	----
Satt251	0.031 [*]	3.8	0.012 [*]	NS	NS	----	----	----	----
Satt14	0.011 [*]	5.3	0.0088 ^{**}	0.033	NS	----	----	----	----

†, *, **, *** Significant at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

NS, Not Significant at the 0.10 level.

Satt251 is significant ($p=0.031$) and shows significant additive effects ($p=0.012$). The maturity covariate is not significant meaning that Satt251 could be a true QTL, which explains 3.8% of the seed oil variation. Marker Satt14 is significant ($p=0.011$) with significant additive and dominant effects ($p=0.0088$ and $p=0.033$, respectively). The maturity covariate is not significant meaning that Satt14 could be a true QTL explaining 5.3 % of the variation in seed oil concentration in this population. Markers Satt540, Satt239, Satt9, Satt50, Satt353, Satt30 were not significant for any values.

The analysis of seed size for the F_2 population at Knoxville is shown in Table 6. Marker Satt540 is significant ($p=0.0027$) and shows significant additive and dominant effects ($p=0.023$ and $p=0.0076$, respectively). Satt540 explained 5.9% of the variation in seed size. The maturity covariate is significant ($p=0.0001$) and Satt540 remains significant, indicating a true QTL. Dominant effects remain significant ($p=0.028$) and the r^2 value reveals that Satt540 explains 21.9% of the variation in seed size. Marker Satt239 is not significant for the marker p -value, but is significant for additive effects ($p=0.059$). The maturity covariate is significant ($p=0.0001$) and the additive effects remain significant ($p=0.073$). The r^2 value indicates that Satt239 is responsible for 20.2% of seed size variation in the covariate model. Satt239 could be a true QTL for seed size. Marker Satt373 is significant ($p=0.024$) and shows significant additive effects ($p=0.0072$). The maturity covariate is significant ($p=0.0001$) and Satt373 remains significant ($p=0.071$) with significant additive effects ($p=0.022$). Satt373 explains 23.4% of the variation in seed size in the covariate model. Marker Satt353 is significant, in the general model, for additive effects ($p=0.099$), but the maturity covariate is significant ($p=0.0001$) and no other values for Satt353 are significant in the maturity-covariate

Table 6. QTL Analysis of Essex x Williams F₂ Soybean Population for Seed Size at Knoxville, TN.

Marker	General Model				Maturity Covariate Model				
	Marker P-value	r ² (%)	Additive P-value	Dominant P-value	Maturity Covariate	Marker P-value	r ² (%)	Additive P-value	Dominant P-value
Satt540	0.0027**	5.9	0.023*	0.0076**	0.0001***	0.082 [†]	21.9	NS	0.028*
Satt52	NS	NS	NS	NS	0.0001***	NS	NS	NS	NS
Satt239	NS	2.2	0.059 [†]	NS	0.0001***	NS	20.2	0.073	NS
Satt373	0.024*	3.8	0.0072**	NS	0.0001***	0.071 [†]	23.4	0.022*	NS
Sctt9	NS	NS	NS	NS	0.0001***	NS	NS	NS	NS
Satt50	NS	NS	NS	NS	0.0001***	NS	NS	NS	NS
Satt353	NS	1.6	0.099 [†]	NS	0.0001***	NS	NS	NS	NS
Satt30	NS	NS	NS	NS	0.0001***	NS	NS	NS	NS
Satt251	NS	NS	NS	NS	0.0001***	NS	NS	NS	NS
Satt14	NS	NS	NS	NS	0.0001***	NS	NS	NS	NS

†, *, **, *** Significant at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

NS, Not Significant at the 0.10 level.

model. This shows that the significance of the QTL in the general model for Satt353 was not really due to seed size, but was due to the way maturity affects the measured expression of that trait. Markers Satt52, Sctt9, Satt50, Satt30, Satt251, and Satt14 were all non-significant as QTLs.

The analysis of plant height for the F_2 population at Knoxville is exhibited in Table 7. Marker Satt540 is significant ($p=0.014$) for plant height and shows additive and dominant effects ($p=0.030$ and $p=0.044$, respectively). The maturity covariate is significant ($p=0.0001$) but no other values for Satt540 are significant in the maturity-covariate model. This indicates that Satt540 is not a true QTL for height and that the significance in the general model was due to maturity's effect on the trait. Satt52 is significant ($p=0.092$) for plant height and shows additive significance ($p=0.033$). In the maturity-covariate model, the maturity covariate is significant ($p=0.0001$) but the marker p -value is not significant. The additive effects are still significant ($p=0.083$) and the r^2 value indicates that Satt52 explains 19.1% of the variation in plant height. Marker Satt373 is significant ($p=0.0001$) for plant height and shows additive and dominant effects ($p=0.0001$ and $p=0.054$, respectively). According to the general model, Satt373 explains 13.5% of plant height variation. The maturity covariate is significant ($p=0.0001$) and the marker p -value and the additive and dominant effects remain significant ($p=0.0001$, $p=0.0001$ and $p=0.024$, respectively). The r^2 value denotes that Satt373 explains 28% of the variation in plant height among progeny. Sctt9 is not significant for any value in the general model, but in the maturity covariate model, the maturity covariate is significant ($p=0.0001$) and the marker p -value is significant ($p=0.055$). Dominant effects are significant ($p=0.051$) and the r^2 value indicates that

Table 7. QTL Analysis of Essex x Williams F₂ Soybean Population for Plant Height at Knoxville, TN.

General Model					Maturity Covariate Model				
Marker	Marker P-value	r ² (%)	Additive P-value	Dominant P-value	Maturity Covariate	Marker P-value	r ² (%)	Additive P-value	Dominant P-value
Satt540	0.014*	4.2	0.030*	0.044*	0.0001***	NS	NS	NS	NS
Satt52	0.092†	2.4	0.033*	NS	0.0001***	NS	19.1	0.083†	NS
Satt239	NS	NS	NS	NS	0.0001***	NS	NS	NS	NS
Satt373	0.0001***	13.5	0.0001***	0.054†	0.0001***	0.0001***	28.0	0.0001***	0.024*
Satt9	NS	NS	NS	NS	0.0001***	0.055†	20.0	NS	0.051†
Satt50	NS	NS	NS	NS	0.0001***	NS	NS	NS	NS
Satt353	NS	NS	NS	NS	0.0001***	NS	NS	NS	NS
Satt30	NS	1.6	0.076†	NS	0.0001***	NS	17.9	0.095†	NS
Satt251	NS	NS	NS	NS	0.0001***	NS	NS	NS	NS
Satt14	NS	1.7	0.080†	NS	0.0001***	NS	NS	NS	NS

†, *, **, *** Significant at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

NS, Not Significant at the 0.10 level.

Sctt9 explains 20% of the variation in plant height, thus Sctt9 may be a true QTL because using a covariate is a statistical way of removing the source of variation due to maturity effects from the residual error term. The result is a tighter test that is better able to find a true QTL marker. Marker Satt30 is significant for additive affects ($p=0.076$) in the general model. In the maturity-covariate model, the maturity covariate is significant ($p=0.0001$) and the additive effects remain significant ($p=0.095$). Satt30 could be a true QTL and explains 17.9% of the variation in plant height. Satt14 is significant for additive effects ($p=0.080$) in the general model, but in the maturity-covariate model, the maturity covariate is significant ($p=0.0001$) and no other p-values for Satt14 are significant. Significance of the QTL in the general model was not really due to plant height, but due to maturity effects on the trait. All other markers, Satt239, Satt50, Satt353, and Satt251, were not significant QTLs.

All the primers and the traits they are significant for among the F_2 population are displayed in Table A.6. Satt540 is significant for maturity, yield, and seed size. In a investigation by Mansur *et al.*, (1996), yield is shown to be very highly correlated with seed number (0.92), in fact, more strongly correlated than with seed weight (-0.32). Yield is also correlated with seed oil (0.60) and seed protein (-0.58). These figures also represent the strong negative correlation between seed protein and seed oil. Several markers were shown to be significant for yield. RFLP markers A6b and R79 explained 6.4% and 12.5%, respectively, of the variation in yield. SSR marker Satt79 explained 6.6% of yield variation (Mansur *et al.*, 1996). In our study, Satt52 is significant for maturity, seed protein, seed oil, and height. Several markers in the study by Mansur *et al.*, (1996), were found to be significant for maturity (stage R8). SSR markers Satt79 and

Satt6 accounted for 19.4% and 12.3%, respectively, of the variation in maturity, and RFLP marker R79 accounted for 18.6% of the variation in maturity (Mansur *et al.*, 1996). In our study, Satt239 is significant for yield and seed size. Several RFLP markers were also found to be significant for seed size, K443 ($R^2=10.8$), A262-3 ($R^2=6.2$), and T155 ($R^2=6.1$). An interesting correlation between seed size and seed number ($R = -0.64$) was also found (Mansur *et al.*, 1996). Satt373 is significant for seed oil, seed size, and height. Mansur *et al.*, (1996), found that three major QTLs account for 48% of the variation in height. Two SSR markers, Satt6 and Satt79 accounted for 31.6% and 18.1%, respectively. Three RFLP markers, R79, A60, and BLT7 accounted for 8%, 7.6%, and 5.5%, respectively, of the height variations (Mansur *et al.*, 1996). In our study, Sctt9 is significant for height. Satt50 is significant for maturity. Satt353 is not significant for any traits. Satt30 is significant for yield and height. Satt251 is significant for seed protein and seed oil. Satt14 is significant for seed oil. Studies by Mansur *et al.*, (1996) indicate that seed protein content and seed oil content are extremely negatively correlated ($R = -0.98$). Two RFLP markers, T155 and A329, and an SSR marker, Satt6, were significant for both seed protein and seed oil. Marker T155 explained 7.3% variation in seed oil content and 9.1% variation in seed protein content. Marker A329 explained 5.2% variation in seed oil and 5.2% variation in seed protein. Marker Satt79 accounted for 8.5% variation in seed oil and 7.8% variation in seed protein (Mansur *et al.*, 1996).

Although traditionally theorized that quantitative traits are controlled by a large number of loci having similar small effects, most of the traits measured in the Minsoy-Noir population used in the Mansur *et al.*, (1996), study were controlled by a few loci with major effects. A number of loci with minor effects were also found varying from as

few as three for lodging to eight for maturity. Some of the major QTLs appeared to effect more than one trait. Three of the larger loci reported on were QTLs associated with yield and yield determining traits.

5.2 F_{4:6} POPULATION QTL ANALYSIS

We wanted to test the same markers against the F_{4:6} generation of the same population to see if using markers to identify traits in an earlier generation are accurate after several generations. We did not expect to be able to identify as many significant dominant QTLs in the later generation because the frequency of heterozygosity decreases by half with every generation of inbreeding (Fehr, 1987). We also did not expect to be able to identify as many significant additive QTLs in the later generation because the QTL associations were based on the F₂ DNA rather than the F_{4:6} DNA, and additive genetic variance increases with inbreeding (Hallauer *et al.*, 1988).

The general model for plant maturity in the F_{4:6} population at Knoxville is shown in Table 8. None of the ten markers were significant for any values. The F₂ population showed two markers, Satt540 and Satt52, significant for a maturity QTL. Previous studies have shown that, for small linkage distances, the proportion of recombination among self-pollinated inbreds is about twice the conventional rate observed in F₂ segregants. Thus deceptive QTL-marker associations due to linkage errors are less likely with recombinant inbred lines than with F₂ or backcross populations (Mansur *et al.*, 1996). Thus, the QTLs for maturity found in the F₂ population may not have been true QTLs.

Table 8. QTL Analysis of Essex x Williams F_{4:6} Soybean Population for Plant Maturity at Knoxville, TN.

Marker	General Model			
	Marker P-value	r ² (%)	Additive P-value	Dominant P-value
Satt540	NS	NS	NS	NS
Satt52	NS	NS	NS	NS
Satt239	NS	NS	NS	NS
Satt373	NS	NS	NS	NS
Sctt9	NS	NS	NS	NS
Satt50	NS	NS	NS	NS
Satt353	NS	NS	NS	NS
Satt30	NS	NS	NS	NS
Satt251	NS	NS	NS	NS
Satt14	NS	NS	NS	NS

NS, not significant at the 0.10 level.

The general and covariate models for plant yield in the F_{4:6} generation at Knoxville is shown in Table 9. Marker Satt52 is significant (0.085) with significant additive effects (0.044). Additive effects refer to the action of genes affecting a trait in a way that each enhances the expression of a trait (Poehlman *et al.*, 1995). The r^2 value indicates that Satt52 explains 3.2% of the variation in yield. The maturity covariate is not significant; therefore Satt52 could be a true QTL.

The general and covariate models for seed protein concentration in the F_{4:6} generation at Knoxville is displayed in Table 10. In the general model, marker Satt239 shows significant dominant effects (0.089) which explain 2.1% of the variation in seed protein concentration. Marker Satt373 is significant (0.030) with significant dominant effects (0.030). Dominance effects are deviations from additivity so that the heterozygote is more like one parent than the other. Dominance is important to breeders of single-cross hybrid crops, such as corn, because both parents are homozygous at all loci resulting in hybrid progeny identical in genotype and uniform in appearance (Poehlman *et al.*, 1995). To breeders of self-pollinated crops, such as soybean, additive effects are desired. Since additive genes each enhance the expression of a trait, breeding self-pollinated crops will result in the reshuffling of genes resulting in many different genotypes among the progeny. The distribution of the progeny will form a typical bell-shaped curve; some of the progeny will have phenotypes superior to both parents and some will have inferior phenotypes (Poehlman *et al.*, 1995). In the general model, Marker Satt373 explains 4.2% of the variation in seed protein concentration. The maturity covariate is significant (0.0001) and so is the marker p-value (0.018), meaning that

Table 9. QTL Analysis of Essex x Williams F_{4:6} Soybean Population for Plant Yield at Knoxville, TN.

Marker	General Model				Maturity Covariate Model				
	Marker P-value	r ² (%)	Additive P-value	Dominant P-value	Maturity Covariate	Marker P-value	r ² (%)	Additive P-value	Dominant P-value
Satt540	NS	NS	NS	NS	NS	----	----	----	----
Satt52	0.085 [†]	3.2	0.044 [*]	NS	NS	----	----	----	----
Satt239	NS	NS	NS	NS	NS	----	----	----	----
Satt373	NS	NS	NS	NS	NS	----	----	----	----
Scrt9	NS	NS	NS	NS	NS	----	----	----	----
Satt50	NS	NS	NS	NS	NS	----	----	----	----
Satt353	NS	NS	NS	NS	NS	----	----	----	----
Satt30	NS	NS	NS	NS	NS	----	----	----	----
Satt251	NS	NS	NS	NS	NS	----	----	----	----
Satt14	NS	NS	NS	NS	0.022 [*]	NS	NS	NS	NS

†, *, **, *** Significant at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

NS, Not Significant at the 0.10 level.

Table 10. QTL Analysis of Essex x Williams F_{4:6} Soybean Population for Seed Protein Concentration at Knoxville, TN.

Marker	General Model				Maturity Covariate Model				
	Marker P-value	r ² (%)	Additive P-value	Dominant P-value	Maturity Covariate	Marker P-value	r ² (%)	Additive P-value	Dominant P-value
Satt540	NS	NS	NS	NS	0.0001***	NS	NS	NS	NS
Satt52	NS	NS	NS	NS	0.0001***	NS	NS	NS	NS
Satt239	NS	2.1	NS	0.089 [†]	0.0001***	NS	NS	NS	NS
Satt373	0.030*	4.2	NS	0.030*	0.0001***	0.018*	17.0	0.048 [†]	0.045 [†]
Scrt9	NS	NS	NS	NS	0.0001***	NS	NS	NS	NS
Satt50	NS	NS	NS	NS	0.0001***	0.071 [†]	15.1	NS	0.068 [†]
Satt353	NS	NS	NS	NS	0.0001***	NS	NS	NS	NS
Satt30	NS	NS	NS	NS	0.0001***	NS	NS	NS	NS
Satt251	NS	NS	NS	NS	0.0001***	NS	NS	NS	NS
Satt14	0.083 [†]	3.3	NS	0.027*	0.0001***	0.011*	51.0	NS	0.012*

†, *, **, *** Significant at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

NS, Not Significant at the 0.10 level.

Satt373 could be a true QTL. There are also significant additive and dominant effects (0.048 and 0.045, respectively). After removing the maturity effects, Satt373 explains 17% of the variation in seed protein concentration. Marker Satt50 does not show any significance in the general model, but the maturity covariate is significant (0.0001). The covariate model shows the marker p-value is significant (0.071) with significant dominant effects (0.068) and that Satt50 explains 15.1% of the variation in seed protein concentration. Marker Satt14 is significant (0.083) in the general model with significant dominant effects (0.027) and it explains 3.3% of the variation in seed protein concentration. The maturity effect is significant (0.0001) and Satt14 remains significant (0.011) in the covariate model, meaning that Satt14 could be a true QTL. The dominance p-value is significant (0.012) and the r^2 value attests that Satt14 is responsible for 51% of the variation in seed protein concentration. All other markers, Satt540, Satt52, Sctt9, Satt353, Satt30, and Sat251 are significant for the maturity covariate, but no other variables.

The general and covariate models for seed oil concentration in the F_{4:6} population at Knoxville is shown in Table 11. Marker Sctt9 is not significant for any values in the general model, but is significant for the maturity covariate (0.0041) and for dominant effects (0.048). Sctt9 could be a true QTL explaining 6.5% of the variation in seed oil concentration. In the general model, marker Satt50 is significant (0.072) with significant additive and dominant effects (0.098 and 0.099, respectively). The maturity covariate is significant (0.009) and the marker remains significant (0.045) meaning that Satt50 could be a true QTL. The additive and dominant effects remain significant (0.076 and 0.069, respectively) and the r^2 value indicates that Satt50 could explain 7.5% of the variation in

Table 11. QTL Analysis of Essex x Williams F_{4:6} Soybean Population for Seed Oil Concentration at Knoxville, TN.

Marker	General Model				Maturity Covariate Model				
	Marker P-value	r ² (%)	Additive P-value	Dominant P-value	Maturity Covariate	Marker P-value	r ² (%)	Additive P-value	Dominant P-value
Satt540	NS	NS	NS	NS	0.0062**	NS	NS	NS	NS
Satt52	NS	NS	NS	NS	0.0092**	NS	NS	NS	NS
Satt239	NS	NS	NS	NS	0.0087**	NS	NS	NS	NS
Satt373	NS	NS	NS	NS	0.0078**	NS	NS	NS	NS
Sc tt9	NS	NS	NS	NS	0.0041**	NS	6.5	NS	0.048*
Satt50	0.072 [†]	3.3	0.098 [†]	0.099 [†]	0.0090**	0.045*	7.5	0.076 [†]	0.069 [†]
Satt353	NS	NS	NS	NS	0.0069**	NS	NS	NS	NS
Satt30	NS	NS	NS	NS	0.0059**	NS	NS	NS	NS
Satt251	NS	NS	NS	NS	0.0085**	NS	NS	NS	NS
Satt14	NS	NS	NS	NS	0.0001***	NS	NS	NS	NS

†, *, **, *** Significant at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

NS, Not Significant at the 0.10 level.

seed oil concentration. All other markers, Satt540, Satt52, Satt239, Satt373, Satt353, Satt30, Satt251, and Satt14 are significant for the maturity covariate, but no other variables.

The analysis of seed size for the $F_{4:6}$ population at Knoxville is displayed in Table 12. In the general model, Satt52 is not significant for any values, but the maturity covariate is significant (0.053) and in the adjusted covariate model, there are significant dominant effects (0.099). Satt52 could be a true QTL explaining 3.8% of the variation in seed size. Marker Satt353 is significant (0.041) in the general model with significant additive and dominant effects (0.085 and 0.072, respectively). The maturity covariate is significant (0.082) and the marker p-value remains significant (0.034) along with the additive and dominant p-values (0.069 and 0.071, respectively) meaning that Satt353 could be a true QTL explaining 5.7% of the variation in seed size. Marker Satt14 is significant (0.0001) in the general model with significant additive and dominant effects (0.06 and 0.0002, respectively). The r^2 value in the general model indicates that Satt14 explains 13.5% of the variation in seed size. In the adjusted covariate model, the maturity covariate is significant (0.046) and the marker p-value remains significant (0.0001) along with the additive and dominant effects (0.034 and 0.0002, respectively) proposing that Satt14 is a true QTL. The r^2 value in the covariate model indicates that Satt14 explains 15.8% of the variation in seed size for this population. Markers Satt540, Satt239, Satt373, Satt9, and Satt251 are all significant for the maturity covariate but no other values. Satt50 and Satt30 are both nonsignificant for all values including the maturity covariate.

Table 12. QTL Analysis of Essex x Williams F_{4:6} Soybean Population for Seed Size at Knoxville, TN.

Marker	General Model				Maturity Covariate Model				
	Marker P-value	r ² (%)	Additive P-value	Dominant P-value	Maturity Covariate	Marker P-value	r ² (%)	Additive P-value	Dominant P-value
Satt540	NS	NS	NS	NS	0.079 [†]	NS	NS	NS	NS
Satt52	NS	NS	NS	NS	0.053 [†]	NS	3.8	NS	0.099 [†]
Satt239	NS	NS	NS	NS	0.054 [†]	NS	NS	NS	NS
Satt373	NS	NS	NS	NS	0.078 [†]	NS	NS	NS	NS
Scct9	NS	NS	NS	NS	0.086 [†]	NS	NS	NS	NS
Satt50	NS	NS	NS	NS	NS	----	----	----	----
Satt353	0.041 [*]	3.9	0.085 [†]	0.072 [†]	0.082 [†]	0.034 [*]	5.7	0.069 [†]	0.071 [†]
Satt30	NS	NS	NS	NS	NS	----	----	----	----
Satt251	NS	NS	NS	NS	0.093 [†]	NS	NS	NS	NS
Satt14	0.0001 ^{**} *	13.5	0.060 [†]	0.0002 ^{***}	0.046 [*]	0.0001 ^{***}	15.8	0.034 [*]	0.0002 ^{***}

†, *, **, *** Significant at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

NS, Not Significant at the 0.10 level.

The general and covariate models for plant height in the $F_{4:6}$ generation at Knoxville are shown in Table 13. Marker Satt239 is significant for additive effects (0.07) in the general model. In the covariate model, the maturity covariate is significant (0.0019) and the additive effects remain significant (0.045) meaning that Satt239 could be a true covariate explaining 7.9% of the variation in plant height for this population. Marker 373 is significant for dominant effects (0.051) in the general model. In the covariate model, the maturity covariate is significant (0.0032) with dominant effects remaining significant (0.07) meaning that Satt373 could be a true covariate explaining 7.7% of the variation in plant height for this population. The remaining markers, Satt540, Satt52, Satt9, Satt50, Satt353, Satt30, Satt251, and Satt14 are significant for the maturity covariate but no other value.

The analysis of plant lodging for the $F_{4:6}$ population at Knoxville is exhibited in Table 14. Marker Satt540 is significant for dominant effects (0.082) in the general model, and the r^2 value indicates that Satt540 explains 2.1% of the variation in plant lodging. The maturity covariate is not significant, therefore marker Satt540 could be a true QTL. Marker Satt14 is significant for additive effects (0.077) with an r^2 value of 2.6%. The maturity covariate is not significant, therefore Satt14 could be a true QTL.

A summary of the markers and the traits significant to the markers for the $F_{4:6}$ Knoxville population are shown in Table A.7. Satt540 is significant for lodging and explains 2.1% of the variation in lodging. Satt52 explains 3.2% of the variation in seed yield and 3.8% of variation in seed size. It seems that seed yield and seed size would be correlated, but according to Mansur *et al.*, (1993) in a study involving $F_{2:5}$ segregating

Table 13. QTL Analysis of Essex x Williams F_{4:6} Soybean Population for Plant Height at Knoxville, TN.

Marker	General Model				Maturity Covariate Model				
	Marker P-value	r ² (%)	Additive P-value	Dominant P-value	Maturity Covariate	Marker P-value	r ² (%)	Additive P-value	Dominant P-value
Satt540	NS	NS	NS	NS	0.0015**	NS	NS	NS	NS
Satt52	NS	NS	NS	NS	0.0021**	NS	NS	NS	NS
Satt239	NS	2.1	0.070 [†]	NS	0.0019**	NS	7.9	0.045*	NS
Satt373	NS	2.6	NS	0.051 [†]	0.0032**	NS	7.7	NS	0.070 [†]
Sctt9	NS	NS	NS	NS	0.0014**	NS	NS	NS	NS
Satt50	NS	NS	NS	NS	0.0026**	NS	NS	NS	NS
Satt353	NS	NS	NS	NS	0.0024**	NS	NS	NS	NS
Satt30	NS	NS	NS	NS	0.0025**	NS	NS	NS	NS
Satt251	NS	NS	NS	NS	0.0025**	NS	NS	NS	NS
Satt14	NS	NS	NS	NS	0.0001***	NS	NS	NS	NS

†, *, **, *** Significant at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

NS, Not Significant at the 0.10 level.

Table 14. QTL Analysis of Essex x Williams F_{4:6} Soybean Population for Plant Lodging at Knoxville, TN.

Marker	General Model				Maturity Covariate Model				
	Marker P-value	r ² (%)	Additive P-value	Dominant P-value	Maturity Covariate	Marker P-value	r ² (%)	Additive P-value	Dominant P-value
Satt540	NS	2.1	NS	0.082 [†]	NS	----	----	----	----
Satt52	NS	NS	NS	NS	NS	----	----	----	----
Satt239	NS	NS	NS	NS	NS	----	----	----	----
Satt373	NS	NS	NS	NS	NS	----	----	----	----
Scrt9	NS	NS	NS	NS	NS	----	----	----	----
Satt50	NS	NS	NS	NS	NS	----	----	----	----
Satt353	NS	NS	NS	NS	NS	----	----	----	----
Satt30	NS	NS	NS	NS	NS	----	----	----	----
Satt251	NS	NS	NS	NS	NS	----	----	----	----
Satt14	NS	2.6	0.077 [†]	NS	NS	----	----	----	----

†, *, **, *** Significant at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

NS, Not Significant at the 0.10 level.

progeny from a cross between 'Minsoy' and 'Noir', seed size was inherited independently of most other traits with the highest correlation being with reproductive stage R1 (flowering). Yield was highly heritable ($H^2 = 0.86$) and was genetically correlated with reproductive stages R1 (flowering), R5 (beginning seed), and R8 (maturity), and plant height. Seed yield has been shown to be only low to moderately heritable in other studies (Burton, 1987). In our study, Satt239 explains 7.9% of the variation in plant height. Satt373 explains 17% variation in seed protein concentration and 7.7% variation in plant height. In the linkage map by Cregan *et al.*, (1999), Satt373 was linked to G1783 by about 22 cM on the Univ. of Utah's linkage group U14. Marker G173 explains 53% of the variation in plant height (Mansur *et al.*, 1993), further evidence that Satt373 could be a true QTL for plant height. Sctt9 explains 6.5% of the variation in seed oil concentration. In the study by Mansur *et al.*, (1993), seed oil was not correlated to any of the other traits in the study. Satt50 was significant ($p=0.071$) for seed protein concentration and explained 15.1% of the variation in seed protein concentration. Seed protein has been shown to be negatively correlated with seed yield (Mansur *et al.*, 1993) and has been recognized as being highly heritable (Burton, 1987) with a value as high as 0.84 (Pantalone *et al.*, 1996). Satt353 and Satt14 explain 5.7% and 15.8%, respectively, of the variation in seed size. Seed size has also been shown to be highly heritable (Burton, 1987) with a value as high as 0.92 (Pantalone *et al.*, 1996). Satt14 also explains 51% of the variation in seed protein concentration and 2.6% of the variation in lodging. Satt30 and Satt251 were not significant for any traits.

The QTL analysis of the Highland Rim Experiment Station (HRES) $F_{4:6}$ population does not include a maturity covariate analysis because although maturity data

were recorded for this population on a field computer, these data were lost and unrecoverable after downloading the data file. We tested the same ten SSR markers, which were identified as polymorphic in the F_2 population, against yield, seed protein concentration, seed oil concentration, seed size, plant height, and lodging.

The general model of the QTL analysis for plant yield at HRES is exhibited in Table 15. Marker Satt353 shows significant additive effects ($p=0.051$). The r^2 value indicates that Satt353 explains 2.4% of the variation in plant yield for this population. No other markers were significant for any traits.

The general model for analysis of seed protein concentration for the $F_{4:6}$ population at HRES is shown in Table 16. Marker Satt353 reveals significant additive effects ($p=0.090$) and explains 2.4% of the variation of seed protein concentration. Marker Satt251 shows significant additive effects ($p=0.077$) for seed protein and explains 2.4% of seed protein variation. No other markers were significant for any measured traits.

The general model for analysis of seed oil concentration for the $F_{4:6}$ population at HRES is shown in Table 17. Marker Satt353 shows significant additive effects ($p=0.067$) and explains 2.5% of the seed oil variation. Marker Satt251 displays significant dominant effects ($p=0.078$). The r^2 value indicates that Satt251 explains 2.6% of the variation in seed oil concentration. No other markers were significant for any traits.

The general model for QTL analysis of seed size for the $F_{4:6}$ population at HRES is displayed in Table 18. Marker Satt50 shows significant additive effects ($p=0.078$) and explains 2.0% of seed size variation. Marker Satt14 is significant ($p=0.016$) and exhibits

Table 15. QTL Analysis of Essex x Williams F_{4:6} Soybean Population for Plant Yield at Highland Rim Experiment Station.

Marker	General Model			
	Marker P-value	r ² (%)	Additive P-value	Dominant P-value
Satt540	NS	NS	NS	NS
Satt52	NS	NS	NS	NS
Satt239	NS	NS	NS	NS
Satt373	NS	NS	NS	NS
Sctt9	NS	NS	NS	NS
Satt50	NS	NS	NS	NS
Satt353	NS	2.4	0.051 [†]	NS
Satt30	NS	NS	NS	NS
Satt251	NS	NS	NS	NS
Satt14	NS	NS	NS	NS

[†], *, **, *** Significant at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

NS, Not Significant at the 0.10 level.

Table 16. QTL Analysis of Essex x Williams F_{4:6} Soybean Population for Seed Protein Concentration at Highland Rim Experiment Station.

Marker	General Model			
	Marker P-value	r ² (%)	Additive P-value	Dominant P-value
Satt540	NS	NS	NS	NS
Satt52	NS	NS	NS	NS
Satt239	NS	NS	NS	NS
Satt373	NS	NS	NS	NS
Sctt9	NS	NS	NS	NS
Satt50	NS	NS	NS	NS
Satt353	NS	2.4	0.090 [†]	NS
Satt30	NS	NS	NS	NS
Satt251	NS	2.4	0.077 [†]	NS
Satt14	NS	NS	NS	NS

†, *, **, *** Significant at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

NS, Not Significant at the 0.10 level.

Table 17. QTL Analysis of Essex x Williams F_{4:6} Soybean Population for Seed Oil Concentration at Highland Rim Experiment Station.

Marker	General Model			
	Marker P-value	r ² (%)	Additive P-value	Dominant P-value
Satt540	NS	NS	NS	NS
Satt52	NS	NS	NS	NS
Satt239	NS	NS	NS	NS
Satt373	NS	NS	NS	NS
Sctt9	NS	NS	NS	NS
Satt50	NS	NS	NS	NS
Satt353	NS	2.5	0.067 [†]	NS
Satt30	NS	NS	NS	NS
Satt251	NS	2.6	NS	0.078 [†]
Satt14	NS	NS	NS	NS

†, *, **, *** Significant at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

NS, Not Significant at the 0.10 level.

Table 18. QTL Analysis of Essex x Williams F_{4:6} Soybean Population for Seed Size at Highland Rim Experiment Station.

Marker	General Model			
	Marker P-value	r ² (%)	Additive P-value	Dominant P-value
Satt540	NS	NS	NS	NS
Satt52	NS	NS	NS	NS
Satt239	NS	NS	NS	NS
Satt373	NS	NS	NS	NS
Sctt9	NS	NS	NS	NS
Satt50	NS	2.0	0.078 [†]	NS
Satt353	NS	NS	NS	NS
Satt30	NS	NS	NS	NS
Satt251	NS	NS	NS	NS
Satt14	0.016 [*]	5.5	NS	0.023 [*]

†, *, **, *** Significant at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

NS, Not Significant at the 0.10 level.

significant dominant effects ($p=0.023$). Satt14 explains 5.5% of the variation in seed size. None of the other markers are significant for any values.

The general model QTL analysis of plant height for the HRES $F_{4:6}$ population is shown in Table 19. Marker Satt239 is significant ($p=0.088$) and shows significant additive effects ($p=0.030$). The r^2 value indicates that Satt239 explains 3.1% of the variation in plant height. Marker Satt373 shows significant dominant effects ($p=0.070$) and explains 2.5% of the plant height variation in this population. No other markers were significant for any values.

The general model for QTL analysis of plant lodging for the HRES $F_{4:6}$ population is displayed in Table 20. Marker Satt251 is significant ($p=0.027$) and shows significant additive effects ($p=0.036$). The r^2 value indicates that Satt251 is responsible for 4.4% of the variation in lodging. No other markers are significant for any traits.

A summary of the markers and the traits significant to the markers for the $F_{4:6}$ Highland Rim population is shown in Table A.8. Markers Satt239 and Satt373 are significant for height in the $F_{4:6}$ population at both Knoxville and Highland Rim. Marker Satt50 was significant for seed size at HRES. Markers Satt353 and Satt251 are significant for seed protein and seed oil. Satt353 is also significant for yield at Highland Rim. Marker Satt14 is significant for seed size in the $F_{4:6}$ Highland Rim population and was very significant for seed protein in the Knoxville $F_{4:6}$ population ($p=0.011$) with a 51% coefficient of determination. Because markers Satt353 and Satt251 are both significant for seed protein and seed oil, it is possible that both traits are correlated or that Satt353 and Satt251 are linked to a QTL for a seed morphological character. In previous

Table 19. QTL Analysis of Essex x Williams F_{4:6} Soybean Population for Plant Height at Highland Rim Experiment Station.

Marker	General Model			
	Marker P-value	r ² (%)	Additive P-value	Dominant P-value
Satt540	NS	NS	NS	NS
Satt52	NS	NS	NS	NS
Satt239	0.088 [†]	3.1	0.030 [*]	NS
Satt373	NS	2.5	NS	0.070 [†]
Sctt9	NS	NS	NS	NS
Satt50	NS	NS	NS	NS
Satt353	NS	NS	NS	NS
Satt30	NS	NS	NS	NS
Satt251	NS	NS	NS	NS
Satt14	NS	NS	NS	NS

†, *, **, *** Significant at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

NS, Not Significant at the 0.10 level.

Table 20. QTL Analysis of Essex x Williams F_{4:6} Soybean Population for Plant Lodging at Highland Rim Experiment Station.

Marker	General Model			
	Marker P-value	r ² (%)	Additive P-value	Dominant P-value
Satt540	NS	NS	NS	NS
Satt52	NS	NS	NS	NS
Satt239	NS	NS	NS	NS
Satt373	NS	NS	NS	NS
Sctt9	NS	NS	NS	NS
Satt50	NS	NS	NS	NS
Satt353	NS	NS	NS	NS
Satt30	NS	NS	NS	NS
Satt251	0.027*	4.4	0.036*	NS
Satt14	NS	2.6	NS	0.066 [†]

[†], *, **, *** Significant at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

NS, Not Significant at the 0.10 level.

studies, seed protein and seed oil have been reported to be negatively correlated (-0.46) (Pantalone *et al.*, 1996).

Marker Satt14 was significant for seed size in the $F_{4:6}$ populations at Knoxville and Highland Rim explaining 15.8 and 5.5%, respectively, of the variation in seed size. There are four markers that are significant for seed size in the Knoxville and Highland Rim $F_{4:6}$ populations, together, explaining between 2.0 and 15.8% variation in seed size. In a study by Maughan *et al.*, (1996), five markers were significant for seed size in an $F_{2:3}$ generation of a *G. max* x *G. soja* cross, and explained from 7.1 to 14.2% of the variation in seed size. The low level of variation explained by these individual markers confirms the quantitative nature of seed size inheritance (Maughan *et al.*, 1996).

Although maturity notes were taken at the Highland Rim Experiment Station, data was lost during downloading, therefore, no maturity covariate analysis was done in the QTL analysis of both locations combined.

The general model for QTL analysis for plant yield for the $F_{4:6}$ population at both locations is shown in Table 21. Marker Satt52 showed significant additive effects ($p=0.086$) and explained 1.8% of the variation in plant yield.

The general model for QTL analysis for seed protein concentration for the $F_{4:6}$ population at both locations is shown in Table 22. Marker Satt373 was significant ($p=0.042$) with significant dominant effects ($p=0.028$). The r^2 value indicates that Satt373 explains 3.8% of the variation in seed protein. Marker Satt251 showed significant additive effects ($p=0.040$) and explained 2.6% of the seed protein variation. Marker Satt14 was significant ($p=0.081$) with significant dominant effects ($p=0.033$). Satt14 explained 3.4% of the variation in seed protein.

Table 21. QTL Analysis of Essex x Williams F_{4:6} Soybean Population for Plant Yield Combined over Knoxville and Springfield, TN.

Marker	General Model			
	Marker P-value	r ² (%)	Additive P-value	Dominant P-value
Satt540	NS	NS	NS	NS
Satt52	NS	1.8	0.086 [†]	NS
Satt239	NS	NS	NS	NS
Satt373	NS	NS	NS	NS
Sctt9	NS	NS	NS	NS
Satt50	NS	NS	NS	NS
Satt353	NS	NS	NS	NS
Satt30	NS	NS	NS	NS
Satt251	NS	NS	NS	NS
Satt14	NS	NS	NS	NS

[†], *, **, *** Significant at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

NS, Not Significant at the 0.10 level.

Table 22. QTL Analysis of Essex x Williams F_{4:6} Soybean Population for Seed Protein Concentration Combined over Knoxville and Springfield, TN.

Marker	General Model			
	Marker P-value	r ² (%)	Additive P-value	Dominant P-value
Satt540	NS	NS	NS	NS
Satt52	NS	NS	NS	NS
Satt239	NS	NS	NS	NS
Satt373	0.042*	3.8	NS	0.028*
Sctt9	NS	NS	NS	NS
Satt50	NS	NS	NS	NS
Satt353	NS	NS	NS	NS
Satt30	NS	NS	NS	NS
Satt251	NS	2.6	0.040*	NS
Satt14	0.081 [†]	3.4	NS	0.033*

[†], *, **, *** Significant at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

NS, Not Significant at the 0.10 level.

The general model for QTL analysis for seed oil concentration for the $F_{4:6}$ population at both locations is displayed in Table 23. Marker Sctt9 was significant for dominant effects ($p=0.053$). The r^2 value indicates that Sctt9 explained 2.7% of the variation in seed oil concentration.

The general model for QTL analysis for seed size for the $F_{4:6}$ population at both locations is shown in Table 24. Marker Satt353 shows significant additive effects ($p=0.093$) and explained 2.5% of seed size variation. Marker Satt14 was highly significant ($p=0.0002$) with highly significant dominant effects ($p=0.0008$) and significant additive effects ($p=0.073$). The r^2 value indicates that Satt14 accounts for 11.3% of the variation in seed size in this population.

The general model for QTL analysis for plant height for the $F_{4:6}$ population for both locations is shown in Table 25. Marker Satt239 showed significant additive effects ($p=0.046$) and explained 2.6% of the variation in plant height. Marker Satt373 was significant for dominant effects ($p=0.071$) and explained 2.4% of plant height variation.

The general model for QTL analysis of plant lodging for the $F_{4:6}$ population for both locations is displayed in Table 26. No markers were significant at the 0.10 level for lodging in the $F_{4:6}$ population at Highland Rim.

A summary of the markers and the traits significant to the markers for both locations, Knoxville and Highland Rim, together is shown in Table A.9. Marker Satt52 was significant for yield. Markers Satt373, Satt251, and Satt14 were significant for seed protein. Marker Sctt9 was significant for seed oil. Markers Satt353 and Satt14 were significant for seed size, and markers Satt239 and Satt373 were significant for plant

Table 23. QTL Analysis of Essex x Williams F_{4:6} Soybean Population for Seed Oil Concentration Combined over Knoxville and Springfield, TN.

Marker	General Model			
	Marker P-value	r ² (%)	Additive P-value	Dominant P-value
Satt540	NS	NS	NS	NS
Satt52	NS	NS	NS	NS
Satt239	NS	NS	NS	NS
Satt373	NS	NS	NS	NS
Sc tt9	NS	2.7	NS	0.053 [†]
Satt50	NS	NS	NS	NS
Satt353	NS	NS	NS	NS
Satt30	NS	NS	NS	NS
Satt251	NS	NS	NS	NS
Satt14	NS	NS	NS	NS

†, *, **, *** Significant at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

NS, Not Significant at the 0.10 level.

Table 24. QTL Analysis of Essex x Williams F_{4:6} Soybean Population for Seed Size Combined over Knoxville and Springfield, TN.

Marker	General Model			
	Marker P-value	r ² (%)	Additive P-value	Dominant P-value
Satt540	NS	NS	NS	NS
Satt52	NS	NS	NS	NS
Satt239	NS	NS	NS	NS
Satt373	NS	NS	NS	NS
Scrt9	NS	NS	NS	NS
Satt50	NS	NS	NS	NS
Satt353	NS	2.5	0.093 [†]	NS
Satt30	NS	NS	NS	NS
Satt251	NS	NS	NS	NS
Satt14	0.0002 ^{***}	11.3	0.073 [†]	0.0008 ^{***}

†, *, **, *** Significant at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

NS, Not Significant at the 0.10 level.

Table 25. QTL Analysis of Essex x Williams F_{4:6} Soybean Population for Plant Height Combined over Knoxville and Springfield, TN.

Marker	General Model			
	Marker P-value	r ² (%)	Additive P-value	Dominant P-value
Satt540	NS	NS	NS	NS
Satt52	NS	NS	NS	NS
Satt239	NS	2.6	0.046*	NS
Satt373	NS	2.4	NS	0.071†
Sctt9	NS	NS	NS	NS
Satt50	NS	NS	NS	NS
Satt353	NS	NS	NS	NS
Satt30	NS	NS	NS	NS
Satt251	NS	NS	NS	NS
Satt14	NS	NS	NS	NS

†, *, **, *** Significant at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

NS, Not Significant at the 0.10 level.

Table 26. QTL Analysis of Essex x Williams F_{4:6} Soybean Population for Plant Lodging Combined over Knoxville and Springfield, TN.

General Model				
Marker	Marker P-value	r² (%)	Additive P-value	Dominant P-value
Satt540	NS	NS	NS	NS
Satt52	NS	NS	NS	NS
Satt239	NS	NS	NS	NS
Satt373	NS	NS	NS	NS
Sctt9	NS	NS	NS	NS
Satt50	NS	NS	NS	NS
Satt353	NS	NS	NS	NS
Satt30	NS	NS	NS	NS
Satt251	NS	NS	NS	NS
Satt14	NS	NS	NS	NS

NS, Not significant at the 0.10 level.

height. No markers were significant for lodging in the QTL analysis of both locations together.

A cross of tomato was made and grown in three different environments, an F_2 population in Davis, CA, the F_3 generation of the same population near Gilroy, CA, and the sibling F_3 generation near Rehovot, Israel (Paterson *et al.*, 1991). The investigation was to see whether QTLs at similar chromosomal locations account for phenotypic variation in different environments. Across the three traits studied, the two California populations shared more QTLs (44%) than either the F_2 or the F_3 California populations did with the Israel population (35 and 19%, respectively) (Paterson *et al.*, 1991).

In our study, 30% of the QTLs were detected across both environments. Two QTLs, Satt239 and Satt373, were found for height at both locations, and marker Satt14 was significant for seed size at both $F_{4:6}$ locations (tables B.6 and B.7). These QTLs were significant enough that they were also picked up in the analysis of both locations together (Table A.9).

Individual QTLs show a range of sensitivities to different environments; some QTLs are detected in all tested environments, whereas some are detected in only one. Classical plant breeders routinely find that genotypes which perform well in one environment are not well suited to other environments, suppressing or lowering the expression of that particular phenotype. Some of these differences among genotypes could be attribute to simply inherited attributes such as susceptibility to particular strains of pest or photoperiod response. QTLs which function consistently over a range of environments are preferred for breeding. By combining several QTLs with different environment specificities into a single genotype, one might illicit an improvement in a

phenotype which is somewhat buffered against the whims of environment (Paterson *et al.*, 1991).

Marker Satt14 shows significant additive and dominant effects for seed size in the F_{4:6} population at Knoxville, TN (Table A.7) and in the analysis of both locations together (Table A.9). This inability to favor either the additive or the dominant model may be limited partly by the number of individuals examined, but may also denote QTLs with partially dominant gene action (Paterson *et al.*, 1991). Indeed, our population of F_{4:6} lines may have sufficient residual heterozygosity that significant dominance genetics effects were detected for this highly heritable trait.

No attempt was made to detect epistasis in this study. All the markers that were significant for a trait were significant for additive, dominant, or both additive and dominant gene actions. The power to detect epistatic effects would be limited by population size. In studies using much larger populations there was still little evidence of epistasis, however, considerable research in quantitative genetics suggests that interactions between QTL alleles at different loci may have a considerable influence on phenotype. Perhaps some of the QTLs that appear to have small effects are dependent upon interaction with other loci, and in smaller populations the optimum allele configurations occur only rarely. Studies of epistasis may be necessary to reveal the genetic basis of subtle phenotypic differences that distinguish successful crop varieties from their less suitable siblings (Paterson *et al.*, 1991). This may provide an area for additional research.

5.3 GENETIC VARIANCE AND HERITABILITY

Heritability is referred to as the effectiveness of selection for a trait depending on the relative importance of genetic and nongenetic factors in the expression of phenotypic differences among genotypes in a population. Heritability can be calculated as the ratio of the genotypic variance (σ_g^2) to the phenotypic variance (σ_{ph}^2). The phenotypic variance can be subdivided into components of variance, σ_e^2 , σ_{ge}^2 , and σ_g^2 , which cause differences in the performance among individuals. The variance component, σ_e^2 , is a measure of difference among genotypes caused by the failure to treat each genotype exactly alike, known as the environmental variance. The symbol σ_{ge}^2 represents differences among phenotypes that are caused by genotype x environment interaction. This is the failure of genotypes to perform the same relative to each other when they are evaluated in different locations and/or years. The genotypic variance, σ_g^2 , is the variation caused by genetic differences among individuals. Genotypic variance is composed of additive, dominant, and epistatic gene effects (Fehr, 1987).

There are two ways that heritability can be expressed, in the broad sense or in the narrow sense. Broad sense heritability is a ratio of the total genetic variance, including additive, dominant, and epistatic effects, to the phenotypic variance. Heritability in the narrow sense is the ratio of the additive genetic variance to the phenotypic variance (Fehr, 1987).

The genetic variance, heritability, mean, and progeny range of each of the measured traits is shown in Table 27. The trait days to maturity showed a genetic variance of 26.6,

Table 27. Table showing the Genetic Variance, Heritability, Mean, and Progeny Range of the F_{4:6} population traits.

	Days to Maturity[†]	Yield (g/plot)	Seed Protein (%DW)	Seed Oil (%DW)	Seed Size (g/100 seed)	Height (cm)	Lodging
Genetic Variance	26.6	0	0.5	0.1	1.4	160.1	0.07
Phenotypic variance	47.5	0	1.3	0.15	1.8	184	0.11
Heritability	0.56	0	0.4	0.65	0.79	0.87	0.61
Mean	115.8	290.1	43.8	19.9	11.9	66.3	1.9
Progeny Range	105-131	172.3-395.3	41.5-46.8	17.8-26.5	8.7-16.2	24.4-98.6	1.3-3.0

[†] Days from planting to R8.

which was then used to calculate the heritability value of 0.56. The mean for this population was 115.8 days to maturity with a progeny range of 105-131 days to maturity. In a study by Mansur *et al.*, (1993), maturity (R8) was highly heritable with an estimate of 0.98. Johnson *et al.*, (1955) estimated heritability over three populations with days to maturity estimated at 0.75.

In our study, yield produced calculated genetic variance and heritability estimates of zero. The mean for yield was 290.1 g/plot with the progeny range being 172.3-395.3 g/plot. Yield distribution for the F_{4:6} population is shown in Figure 3. According to a study by Pantalone *et al.*, (1996), seed yield has been shown to have a low to moderate heritability (0.36). In other studies yield has been shown to be highly heritable at 86%, and also highly correlated with the reproductive stages R1 (0.70), R5 (0.80), and R8 (0.80) and plant height (0.68), all of which are traits known to have high heritability (Mansur *et al.*, 1993). In another study yield, measured over 2-3 locations and 2 years, had a genetic variance of 1.72 and heritability of 25.2% (Johnson *et al.*, 1955).

Seed protein concentration had a genetic variance estimate of 0.5 and heritability of 0.4. The mean for seed protein in this population was 43.8% dry weight of seed (DW) with a progeny range of 41.5-46.8% DW. Seed protein distribution is shown in Figure 4. Seed protein has been shown to be highly heritable in several studies with values of 0.84 (Pantalone *et al.*, 1996), 0.83 (Mansur *et al.*, 1993), and 0.82 (Johnson *et al.*, 1955). Also in one population studied by Johnson *et al.*, (1955), seed protein has shown a lower heritability of 0.39.

In this study, seed oil concentration had a genetic variance of 0.1 and a heritability estimate of 0.65. The mean for this population was 19.9% DW with a progeny range of

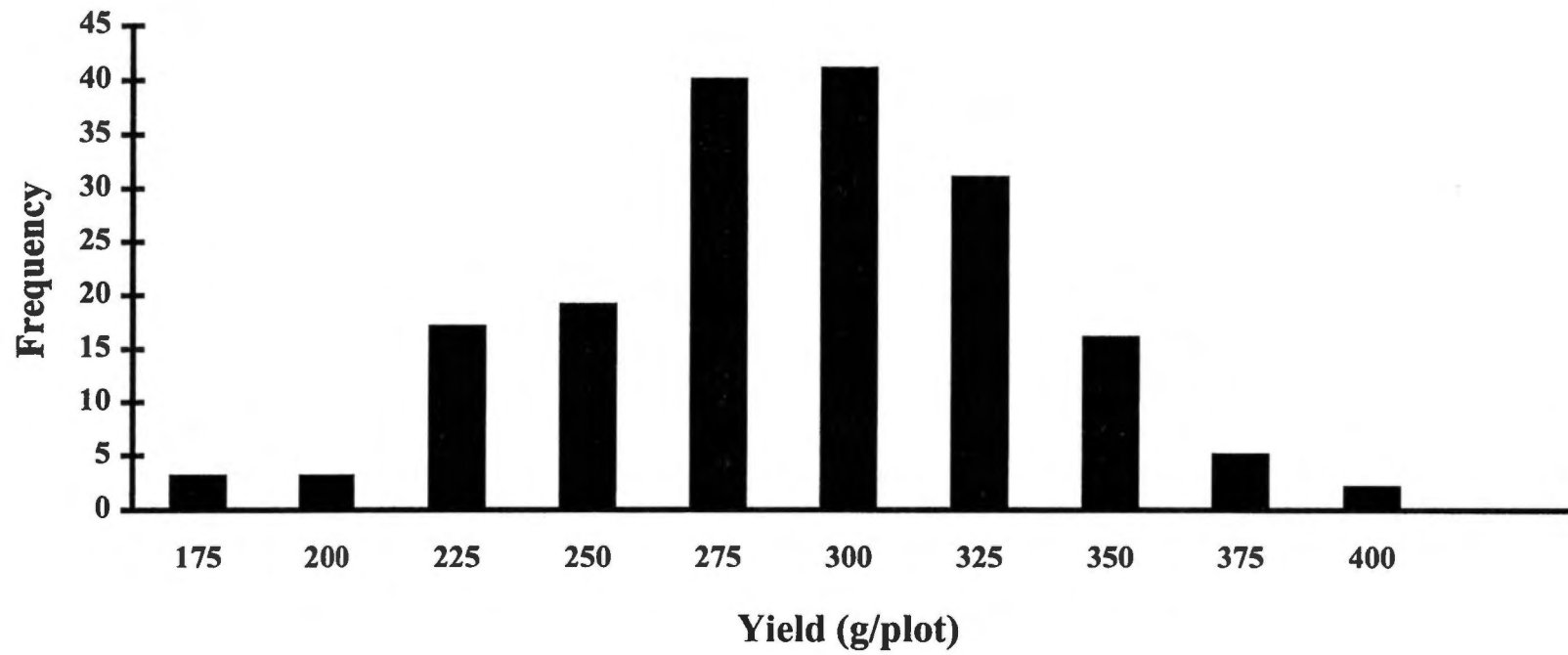


Figure 3. Frequency Distribution for Mean Plot Yield of Essex x Williams F_{4:6} Population over two Locations

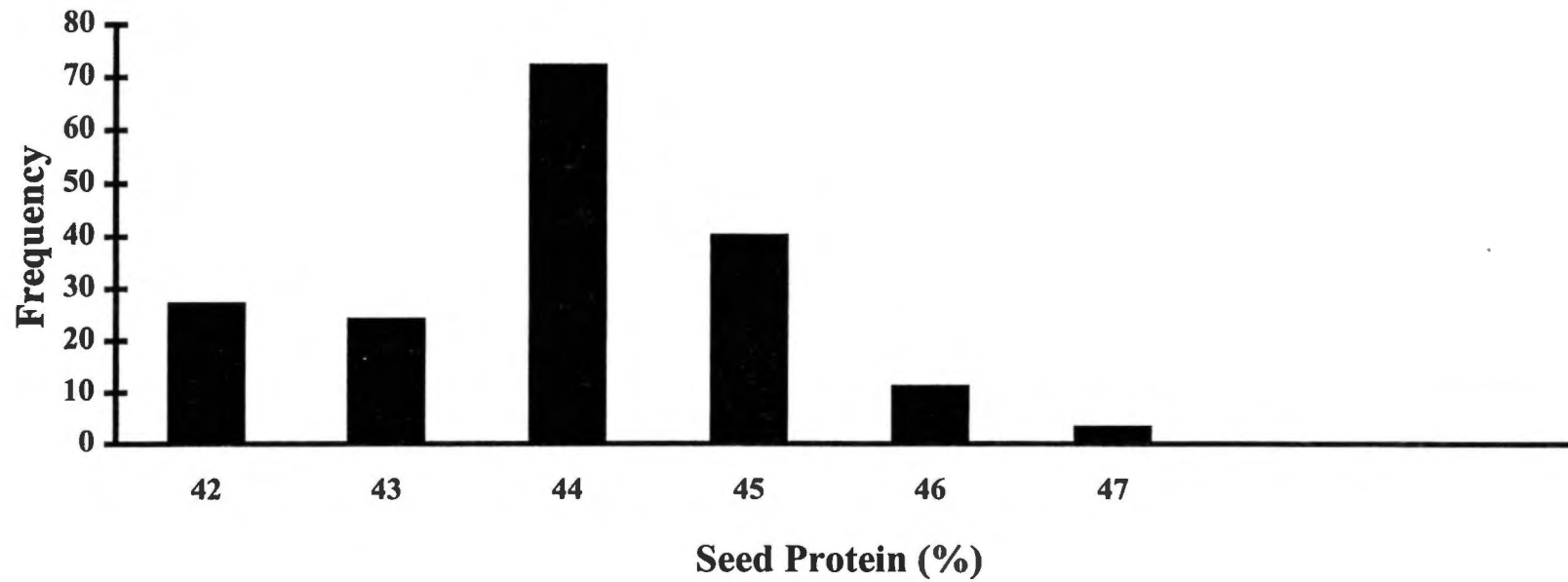


Figure 4. Frequency Distribution for Mean Seed Protein Concentration of Essex x Williams F_{4:6} Population over two Locations

17.8-26.5% DW. Seed oil has shown high heritability in several studies: 0.83 (Pantalone *et al.*, 1996), 0.88 (Mansur *et al.*, 1993), and 0.78 (Johnson *et al.*, 1955). variance has been estimated as 0.159 and 0.289 in other studies (Johnson *et al.*, 1955).

Seed size had a genetic variance of 1.4 and a heritability estimate of 0.79 in this study. The mean was 11.9 g/100 seed with a progeny range 8.7-16.2 g/100 seed. In other studies, seed size heritability has been estimated at 0.92 (Pantalone *et al.*, 1996), 0.94 (Mansur *et al.*, 1993), 0.68, and 0.92 (Johnson *et al.*, 1955). Genetic variance has been estimated at 0.53 and 4.06 (Johnson *et al.*, 1955).

Plant height had a genetic variance of 160.1 and a heritability estimate of 0.87 in this study. The mean was 66.3 cm with a progeny range of 24.4-98.6 cm. Other studies have estimated height heritability at 0.93 (Mansur *et al.*, 1993), 0.61, and 0.81 (Johnson *et al.*, 1955). Genetic variance estimates have been measured at 7.19 and 9.34 (Johnson *et al.*, 1955).

In this study, plant lodging had a genetic variance estimate of 0.07 and a heritability estimate of 0.61. The mean lodging score was 1.9 with a progeny range of lodging scores 1.3-3.0. The heritability estimates of lodging in other studies has ranged from 0.34 (Johnson *et al.*, 1955) to 0.76 (Mansur *et al.*, 1993).

Estimates of genetic variance and heritability can be useful in several areas of plant breeding: assembly or creation of a pool of variable germplasm, selection of superior individuals from the pool, and utilization of the selected individuals to create a superior variety. Detection of significant genetic variability indicates that genetic variation exists in the germplasm pool but says nothing about the range of genetic variability within a particular population. A heritability estimate from such a test provides information on

the relative magnitudes of genetic and environmental variation that exists in the germplasm pool (Dudley *et al.*, 1969).

6. CONCLUDING REMARKS AND FURTHER RESEARCH

For future study, the population used in this study can be tested with more SSR markers and the PCR products from Essex and Williams can be sequenced and compared with known sequences in one of many databases. A bulk segregant analysis can be done to determine qualitative genes for flower color, pubescence color and growth habit. Mapmaker can be utilized to map the significant markers and QTLs, then compared with existing soybean genome maps.

Plant breeders have been very successful in using phenotypic observations or measurements, more recently backed up by sophisticated statistical methods to select superior individuals in breeding populations. Usually, large numbers of individuals and many replications are required to achieve statistical validity (Rafalski *et al.*, 1994). Since some of the desired and economically important traits are quantitative traits and are highly influenced by the environment, the visual selection or selection based on phenotypic estimations may not be effective. In other cases, a qualitative trait like disease susceptibility may not express if suitable environmental conditions are not available. Contrasting traits may not be distinguishable at the seedling stage, making it necessary to grow the population to the adult stage before exercising selection. Raising large populations to the mature stage for selection requires effort and expense, and therefore a breeder would like to execute selection at the seedling stage if feasible. These difficulties could be largely overcome if molecular marker-assisted plant breeding could be utilized in a cost-effective, efficient manner (Gupta *et al.*, 1996).

The positive and negative alleles for the additive QTLs that were consistent over different environments are shown in Table A.10. Consistency over environments is of interest to some breeders in order to make a soybean line attractive for a larger area. Being able to choose a breeding line based on selection of certain alleles of a trait will greatly facilitate the selection process.

Molecular breeding is becoming more practical as more and better technology emerges. Great progress has been made in the construction of genetic maps in many plant species, including all major crops. Most of these maps are constructed with RFLP markers, and many new maps are being constructed with both RADP and microsatellite markers (Rafalski *et al.*, 1994) (Cregan *et al.*, 1999).

The Essex x Williams population evaluated in our experiments is unique because of its southern and northern ancestry. This cross has contributed to many important commercial and public soybean cultivars. Continued molecular genetics and genomics research with the Essex x Williams population may further help to develop applied, molecular marker-assisted approaches for the benefit of soybean breeders.

LITERATURE CITED

LITERATURE CITED

- Allen, F.L. 1994. Usefulness of Plant Genome Mapping to Plant Breeding. In "Plant Genome Analysis" (P.M. Gresshoff, Ed.), pp. 11-18. CRC Press, Boca Raton.
- Akkaya, M.S., P. B. Cregan, A. Bhagwat, U. Lavi. 1993. DNA Fingerprinting of Soybean Cultivars using Simple Sequence Repeat (SSR) DNA Markers. In "DNA Fingerprinting Workshop" (G. Caetano-Anolles, Coordinator), pp. 1-7. Racheff Chair in PMG, Univ. of Tenn. Knoxville.
- Akkaya, M.S., R. C. Shoemaker, J. E. Specht, A. A. Bhagwat, P. B. Cregan. 1995. Integration of Simple Sequence Repeat DNA Markers into a Soybean Linkage Map. *Crop Sci.* 35:1439-1445
- Bassam, B. J., G. Caetano-Anolles, P. M. Gresshoff. 1991. Fast and Sensitive Silver Staining of DNA in Polyacrylamide Gels. *Anal. Biochem.* 196:80-83.
- Bernard, R. L., D. A. Lindahl. 1972. Registration of Williams Soybean. *Crop Sci.* 12:716.
- Brim, C. A., C. W. Stuber, 1973. Application of genetic male sterility to recurrent selection schemes in soybeans. *Crop Sci.* 13:528-830.
- Brummer, E.C., A. D. Nickell, J. R. Wilcox, R. C. Shoemaker. 1995. Mapping the *Fan* Locus Controlling Linolenic Acid Content in Soybean Oil. *J. Hered.* 86:245-247.
- Burton, J. W. 1987. Quantitative genetics: results relevant to soybean breeding. p. 211-247. In R. Shibles (ed.) Soybeans: Improvement, production, and uses, 2nd ed. Agron. Monogr. 16. ASA, CSSA, SSSA, Madison, WI.
- Carter, T.E. Jr., Z. Gizlice, J. W. Burton. 1993 Coefficient of parentage and genetic-proximity estimates for 258 North American cultivars released by public agencies between 1945 and 1988. *USDA Tech. Bull.* 1814. U.S. Gov. Print Office, Washington D.C.
- Cober, E.R., H. D. Voldeng, J. A. Fregeau-Reid. 1997. Heritability of Seed Shape and Seed Size in Soybean. *Crop Sci.* 37: 1767-1769.
- Cregan, P. B., M. S. Akkaya, A. A. Bhagwat, U. Lavi, J. Rongwen. 1994. Length Polymorphisms of Simple Sequence Repeat (SSR) DNA as Molecular Markers in Plants. In "Plant Genome Analysis" (P.M. Gresshoff, Ed.), pp. 47-56. CRC Press, Boca Raton.

- Cregan, P. B., T. Jarvik, A. L. Bush, R. C. Shoemaker, K. G. Lark, A. L. Kahler, N. Kaya, T. T. VanToai, D. G. Lohnes, J. Chung, J. E. Specht. 1999. An Integrated Genetic Linkage Map of the Soybean Genome. *Crop Sci.* 39:1464-1490.
- Dellaporta, S. J., J. Wood, J. B. Hicks. 1983. A Plant DNA mini-preparation, Version II. *Plant Mol. Biol.* 1:19-21.
- Diers, B.W., P. Keim, W. R. Fehr, R. C. Shoemaker. 1992. RFLP analysis of soybean seed protein and oil content. *Theor. Appl. Genet.* 83:608-612.
- Dudley, J. W., R. H. Moll. 1969. Interpretation and Uses of Estimates of Heritability and Genetic Variances in Plant Breeding. *Crop Sci.* 9:257-262.
- Falconer, D. S. 1989. "Introduction to Quantitative Genetics". 3rd ed. Longman Scientific and Technical. England. pp. 117-121
- Fehr, W. 1987. "Principles of Cultivar Development: Crop Species" Macmillan Publishing Company, New York. pp. 540-544.
- Fehr, W. 1987. "Principles of Cultivar Development: Theory and Technique" Macmillan Publishing Company, New York
- Grattapaglia, D. 1994. Genetic mapping of quantitatively inherited economically important traits in eucalyptus. North Carolina State University, Raleigh. Ph.D dissertation.
- Gupta, P. K., H. S. Balyan, P. C. Sharma, B. Ramesh. 1996. Microsatellites in plants: A new class of molecular markers. *Curr. Sci.* 70:45-54.
- Hallauer, A. R., J. B. Miranda. 1988. Quantitative genetics in maize breeding. Iowa State University Press, Ames, IA.
- Johnson, H. W., H. F. Robinson, R. E. Comstock. 1955. Estimates of genetic and environmental variability in soybeans. *Agron. J.* 47:314-318.
- Lark, K. G., J. Orf, L. M. Mansur. 1994. Epistatic expression of quantitative trait loci (QTL) in soybean [*Glycine max* (L.) Merr.] determined by QTL association with RFLP alleles. *Theor. Appl. Genet.* 88:486-789.
- Lee, S. H., M. A. Bailey, A. R. Mian, T. E. Carter Jr., D. A. Ashley, R. S. Hussey, W. A. Parrott, and H. R. Boerma. 1996. Molecular Markers Associated with Soybean Plant Height, Lodging, and Maturity across Locations. *Crop Sci.* 36:728-735.

- Mansur, L. M., K. G. Lark, H. Kross, A. Oliviera. 1993. Interval mapping of quantitative trait loci for reproductive, morphological, and seed traits of soybean (*Glycine max* L.). *Theor. Appl. Genet.* 86:907-913.
- Mansur, L. M., J. H. Orf, K. Chase, T. Jarvik, P. B. Cregan, K. G. Lark. 1996. Genetic Mapping of Agronomic Traits Using Recombinant Inbred Lines of Soybean. *Crop Sci.* 36:1327-1336.
- Mansur, L. M., J. Orf, K. G. Lark. 1993. Determining the linkage of quantitative trait loci to RFLP markers using extreme phenotypes of recombinant inbreds of soybean (*Glycine max* L. Merr.). *Theor. Appl. Genet.* 86:914-918.
- Maughan, P. J., M. A. Saghai Maroof, G. R. Buss. 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.
- McIntosh, M. S. 1983. Analysis of Combined Experiments. *Agron. J.* 75:153-155.
- Mian, A. R., M. A. Bailey, J. P. Tamulonis, E. R. Shipe, T. E. Carter Jr., W. A. Parrott, D. A. Ashley, R. S. Hussey, H. R. Boerma. 1996. Molecular markers associated with seed weight in two soybean populations. *Theor. Appl. Genet.* 93:1011-1016.
- Morgante, M., A. Rafalski, P. Biddle, S. Tingey, A.M. Olivieri. 1994. Genetic mapping and variability of seven soybean simple sequence repeat loci. *Genome.* 37:763-769.
- Pantalone, V. R., J. W. Burton, T.E. Carter Jr. 1996. Soybean Fibrous Root Heritability and Genotypic Correlations with Agronomic and Seed Quality Traits. *Crop Sci.* 36:1120-1125.
- 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.* 2127:181-197.
- Poehlman, J.M., D. A. Sleper. 1995. "Breeding Field Crops", 4th edition. Iowa State University Press. Ames, Iowa.
- Rafalski, J. A., M. K. Hanafey, S. V. Tingey, J. G. K. Williams. 1994. Technology for Molecular Breeding: RAPD Markers, Microsatellites and Machines. In "Plant Genome Analysis" (P.M. Gresshoff, Ed.), pp. 19-27. CRC Press, Boca Raton.

- Rawlings, J. O. 1988. "Applied Regression Analysis, A Research Tool". Wadsworth and Brooks/Cole Advanced Books and Software. Pacific Grove, California. p.213.
- Rongwen, J., M. S. Akkaya, A. A. Bhagwat, U. Lavi, P. B. Cregan. 1995. The use of microsatellite DNA markers for soybean genotype identification. *Theor. Appl. Genet.* 90:43-48.
- Shoemaker, R. C., L. L. Lorenzen, B. W. Diers, T. C. Olson. 1994. Genome Mapping and Agriculture. In "Plant Genome Analysis" (P.M. Gresshoff, Ed.), pp. 1-10. CRC Press, Boca Raton.
- Smith, T. J., H. M. Camper. 1973. Registration of Essex Soybean. *Crop Sci.* 13: 495.
- Trigiano, R. N., G. Caetano-Anolles. 1998. Laboratory Exercises on DNA Amplification Fingerprinting for Evaluating the Molecular Diversity of Horticultural Species. *Hort Technology.* 8:413-423.
- Ustun, A. 1999. Variations for some Quantitative Traits and Molecular Markers in some Ancestors and Cultivars of Soybean, and Evaluation of Yield Increase by Breeding in the Southern U.S. University of Tennessee at Knoxville. Ph.D Dissertation.
- Vogel, J. M., P. A. Scolnik. 1997. Direct amplification from microsatellites: detection of simple sequence repeat-based polymorphisms without cloning. In "DNA Markers: Protocols Applications and Overviews" (G. Caetano-Anolles, P.M. Gresshoff, Eds.), pp. 133-150. Wiley-VCH, New York.
- Weising, K., P. Winter, B. Huttel, G. Kahl. 1998. Microsatellite Markers for Molecular Breeding. *J. Crop Prod.* 1:113-143.
- Young, N. D. 1996. QTL Mapping and Quantitative Disease Resistance in Plants. *Annu. Rev. Phytopathol.* 34:479-501.
- Yu, Y. G., M. A. Saghai Maroof, G. R. Buss, P. J. Maughan, S. A. Tolin. 1993. RFLP and Microsatellite Mapping of a Gene for Soybean Mosaic Virus Resistance. *Phytopathology.* 84:60-64.
- www.ars.grin.gov
- www.hort.purdue.edu/newcrop/proceedings1990/V1-159.html
- www.omnitechintl.com/soy_program/overview.html
- www.soymeal.org/ksmith1.html

APPENDIX

Table A.1. Partial List of Soybean Cultivars with Essex x Williams in their lineage.

Variety	Reference	Maturity Group	Parents	W x E Ancestor	Favorable Attributes	Year Released
CF461	Crop Science	IV	Asgrow 3127 x Maple Isle	A3127	Superiority in seed yield	1995
Conrad	Crop Science	II	A3127 x Tri-Valley Charger	A3127	Superiority in seed yield	1988
Defiance	Crop Science	III	HM8778 x Asgrow A3733	A3127	High yield and resistance to Phytophthora rot	1996
Delsoy 4900	Crop Science	IV	Asgrow A5424 x Mack	A5424	High yield and resistance to SCN	1989
Dunbar	Crop Science	III	Platte x A3127	A3127	High yield	1991
Flyer	Crop Science	IV	A3127 x Williams 82(L24), Rpsl-k	A3127	High yield and resistance to Phytophthora rot	1988
GR8836	Crop Science	III	A3127 x Williams 82(L24), Rpsl-k	A3127	High yield and resistance to Phytophthora rot	1987
GR8936	Crop Science	III	A3127 x Williams 82(L24), Rpsl-k	A3127	High yield and resistance to Phytophthora rot	1988
KS3494	Crop Science	III	Harper x A3127	A3127	High yield	1993

Table A.1 cont'd. Partial List of Soybean Cultivars with Essex x Williams in their lineage.

Variety	Reference	Maturity Group	Parents	W x E Ancestor	Favorable Attributes	Year Released
LN89-3615	Crop Science	IV	Hobbit 87 x Asgrow A3205	A3127	High yield and resistance to Phytophthora rot	1996
LN92-11008	Crop Science	III	Jack x A3205	A3127	Resistance to SCN and higher yield than other SCN-resistant varieties	1996
Omaha	Crop Science	IV	A3733 x Resnick	A3127	High yield and resistance to Phytophthora rot	1996
Pana	Crop Science	III	Jack x A3205	A3127	Resistance to SCN and higher yield than other SCN-resistant varieties	1997
Pennyrile	Crop Science	IV	Williams x Essex	----	High yield	1987
Piatt	Crop Science	III	F ₄ LN78-257 x A3127	A3127	High yield and semideterminate growth type	1993
Resnick	Crop Science	III	A3127 x Williams 82(L24), Rpsl-k	A3127	High yield and resistance to Phytophthora rot	1987

Table A.1 cont'd. Partial List of Soybean Cultivars with Essex x Williams in their lineage.

Variety	Reference	Maturity Group	Parents	W x E Ancestor	Favorable Attributes	Year Released
Savoy	Crop Science	II	Burlison x A3733	A3127	High yield and resistance to Phytophthora rot	1996
Stressland	Crop Science	IV	HC80-1946 X A3127	A3127	Tall, indeterminate, drought tolerant and vigorous vegetative growth. Very adapted to stress-prone environs	1994
Thorne	Crop Science	III	F ₅ A80-344003 x A3127BC ₃ F ₂ -1	A3127	Superior yield and disease resistance	1992
Cisne	Crop Science	IV	Burlison x A3733	A3127	High yield and resistance to Phytophthora rot	1995
Iroquois	Crop Science	III	F ₄ LN81-1029 x Asgrow A2943	A3127	High yield and resistance to Phytophthora rot	1995
LN90-4524	Crop Science	III	F ₄ LN86-4668 x A3733	A3127	Resistance to Races 3, 4, 5 and 14 to of SCN and higher yield	1995

Literature Cited for Table A.1

- Anand, S.C. 1991. Registration of 'Delsoy 4900' Soybean. *Crop Sci.* 31:849-850.
- Cooper, R.L., R.J. Martin, S.K. St. Martin, A.J. Calip-DuBois, R.J. Fioritto, A.F. Schmitthenner. 1999. Registration of 'Stressland' Soybean. *Crop Sci.* 39:590-591.
- Fehr, W.R., S.R. Cianzio, B.K. Voss, S.P. Schultz. 1989. Registration of 'Conrad' Soybean. *Crop Sci.* 29:830.
- Graef, G.L., J.E. Specht, L.L. Korte, D.M. White. 1992. Registration of 'Dunbar' Soybean. *Crop Sci.* 32:497.
- McBlain, B.A., R.J. Fioritto, S.K. St. Martin, A.J. Calip-DuBois, A.F. Schmitthenner, R.L. Cooper, R.J. Martin. 1990. Registration of 'Resnick' Soybean. *Crop Sci.* 30:424-425.
- McBlain, B.A., R.J. Fioritto, S.K. St. Martin, A.J. Calip-DuBois, A.F. Schmitthenner, R.L. Cooper, R.J. Martin. 1990. Registration of 'Flyer' Soybean. *Crop Sci.* 30:425.
- McBlain, B.A., R.J. Fioritto, S.K. St. Martin, A.J. Calip-DuBois, A.F. Schmitthenner, R.L. Cooper, R.J. Martin. 1990. Registration of 'GR8836' Soybean. *Crop Sci.* 30:425-426.
- McBlain, B.A., R.J. Fioritto, S.K. St. Martin, A.J. Calip-DuBois, A.F. Schmitthenner, R.L. Cooper, R.J. Martin. 1990. Registration of 'GR8936' Soybean. *Crop Sci.* 30:426-427.
- McBlain, B.A., R.J. Fioritto, S.K. St. Martin, A.J. Calip-DuBois, A.F. Schmitthenner, R.L. Cooper, R.J. Martin. 1993. Registration of 'Thorne' Soybean. *Crop Sci.* 33:1406.
- Nickell, C.D., D.J. Thomas, K. Frey. 1994. Registration of 'Piatt' Soybean. *Crop Sci.* 34:1128-1129.
- Nickell, C.D., G.R. Noel, J.E. Tharp, T.R. Cary, D.J. Thomas. 1996. Registration of 'LN90-4524' Soybean. *Crop Sci.* 36:1411.
- Nickell, C.D., D.J. Thomas, T.R. Cary, E.O. Johnson, D. Gay. 1996. Registration of 'Iroquois' Soybean. *Crop Sci.* 36:1411-1412.
- Nickell, C.D., D.J. Thomas, T.R. Cary, P.A. Stephens. 1996. Registration of 'Cisne' Soybean. *Crop Sci.* 36:1410.

- Nickell, C.D., M.S. Bachman, D.J. Thomas, T.R. Cary. 1998. Registration of 'Omaha' Soybean. *Crop Sci.* 38:547.
- Nickell, C.D., G.R. Noel, T.R. Cary, D.J. Thomas. 1998. Registration of 'Pana' Soybean. *Crop Sci.* 38:1398-1399.
- Nickell, C.D., G.R. Noel, J.E. Tharp, T.R. Cary, D.J. Thomas. 1998. Registration of 'LN92-11008' Soybean. *Crop Sci.* 38:546.
- Nickell, C.D., D.J. Thomas, T.R. Cary, J.M. Hegstad. 1998. Registration of 'LN89-3615' Soybean. *Crop Sci.* 38:548
- Nickell, C.D., D.J. Thomas, T.R. Cary, D.E. Kyle, J.M. Hegstad. 1998. Registration of 'Savoy' Soybean. *Crop Sci.* 38:547.
- Pfeiffer, T.W., J.H. Orf. 1988. Registration of 'Pennyrile' Soybean. *Crop Sci.* 28:719-720.
- Pfeiffer, T.W., L.C. Harris, D.L. Pilcher, J.M. Wood. 1996. Registration of 'CF461' Soybean. *Crop Sci.* 36:811.
- Schapaugh, W.T. 1998. Registration of 'KS3494' Soybean. *Crop Sci.* 38:891.
- St. Martin, S.K., A.J. Calip-DuBois, R.J. Fioritto, A.F. Schmitthenner, R.L. Cooper, R.J. Martin. 1997. Registration of 'Defiance' Soybean. *Crop Sci.* 37:1981.

Table A.2. Cultivar Descriptions of Essex and Williams.

Trait	Essex	Williams
Maturity Group	V	III
Yield (Mg/ha)	2.67	4.02
Flower Color	Purple	White
Plant Height (cm)	116	94
Growth Habit	Determinate	Indeterminate
Pubescence Color	Gray	Tawny
Seed Oil Conc. (% DW)	19.6	19.8
Seed Protein Conc. (%DW)	42.5	45.6

<http://www.ars.grin.gov>

Table A.3. List of Screened Markers Showing Fragment Size for Markers that are Polymorphic Between Cultivars Essex and Williams.

SSR Marker	Polymorphic	Essex fragment length (bp)	Williams fragment length (bp)
Sat1	No	----	----
Satt100	No	----	----
Satt127	No	----	----
Satt132	No	----	----
Satt134	No	----	----
Satt14	Yes	205	183
Satt144	No	----	----
Satt173	No	----	----
Satt180	No	----	----
Satt192	No	----	----
Satt194	No	----	----
Satt197	No	----	----
Satt2	No	----	----
Satt20	No	----	----
Satt202	No	----	----
Satt22	No	----	----
Satt222	No	----	----
Satt225	No	----	----
Satt236	No	----	----
Satt239	Yes	208	205
Satt251	Yes	216	226
Satt253	No	----	----
Satt276	No	----	----
Satt289	No	----	----
Satt30	Yes	168	176
Satt307	No	----	----
Satt31	No	----	----
Satt319	No	----	----
Satt32	No	----	----
Satt353	Yes	174	165
Satt368	No	----	----
Satt371	No	----	----
Satt373	Yes	233	261
Satt380	No	----	----
Satt383	No	----	----

Table A.3 cont'd. List of Screened Markers Showing Fragment Size for Markers that are Polymorphic Between Cultivars Essex and Williams.

SSR Marker	Polymorphic	Essex fragment length (bp)	Williams fragment length (bp)
Satt396	No	----	----
Satt409	No	----	----
Satt426	No	----	----
Satt45	No	----	----
Satt455	No	----	----
Satt460	No	----	----
Satt469	No	----	----
Satt482	No	----	----
Satt489	Yes	----	----
Satt49	No	----	----
Satt496	No	----	----
Satt50	Yes	172	177
Satt509	No	----	----
Satt513	No	----	----
Satt52	Yes	241	246
Satt540	Yes	170	160
Satt554	No	----	----
Satt588	No	----	----
Satt6	No	----	----
Satt603	Yes	----	----
Satt89	No	----	----
Sctt9	Yes	105	111

Table A.3. Forward and Reverse Primer Sequences of Polymorphic SSR Markers.

SSR Marker	Forward primer sequence	Reverse primer sequence
Satt540	CTGGCGAATCAACTTTGTAAC	CCGTGATTGCGAAGAGGATATT
Satt52	GAATAAAATTAGGATAAGTGATAAG	ACAGAAAAAAGAAAAATGTCA
Satt239	GCGCCAAAAAATGAATCACAAT	GCGAACACAATCAACATCCTTGAAC
Satt373	TCCGCGAGATAAATTCGTAAAAT	GGCCAGATACCCAAGTTGTACTTGT
Sctt9	TGGTAAAGGAGGAACTT	AGAATGTGCTGATGACA
Satt50	GAACGTCCCAACAATG	GGCCTTGTAACATAT
Satt353	CATACACGCATTGCCTTTCCTGAA	GCGAATGGGAATGCCTTCTTATTCTA
Satt30	AAAAAGTGAACCAAGCC	TCTTAAATCTTATGTTGATGC
Satt251	CCTCCACCCCCTTCCCACCCAAAA	GGTGATATCGCGCTAAAATTA
Satt14	TTTTTTTCTCTTCATGTTTG	CCCGTATAGCAACCAA

<http://129.186.26.94> (Soybase homepage)

Table A.5 Expected Mean Squares for Evaluating a Random Set of Genotypes over Random Environments.

Source of Variation	df	MS	EMS
Environment	e-1	M ₁	$\sigma^2 + r \sigma_{GE}^2 + g \sigma_{R(E)}^2 + rg \sigma_E^2$
Rep (Env)	e(r-1)	M ₂	$\sigma^2 + g \sigma_{R(E)}^2$
Genotype	g-1	M ₃	$\sigma^2 + r \sigma_{GE}^2 + re \sigma_G^2$
Gen x Env	(g-1)(e-1)	M ₄	$\sigma^2 + r \sigma_{GE}^2$
Error	e(r-1)(g-1)	M ₅	σ^2

Table A.6 Summarized List of Traits per Marker for the F₂ Population at Knoxville, TN.

Marker	Trait	Marker p-value	r ² (%)	Additive p-value	Dominant p-value
Satt540	Maturity	0.0001 ^{***}	10.6	0.0001 ^{***}	0.054 [†]
	Yield	0.086 [†]	13.3	NS	0.098 [†]
	Seed Size	0.082 [†]	21.9	NS	0.028 [*]
Satt52	Maturity	NS	2.0	NS	0.059 [†]
	Seed Protein	0.064 [†]	3.0	NS	0.021 [*]
	Seed Oil	0.062 [†]	3.0	NS	0.021 [*]
	Height	NS	19.1	0.083 [†]	NS
Satt239	Yield	0.087 [†]	13.0	0.027 [*]	NS
	Seed Size	NS	20.2	0.073 [†]	NS
Satt373	Seed Oil	NS	1.8	0.077 [†]	NS
	Seed Size	0.071 [†]	23.4	0.022 [*]	NS
	Height	0.0001 ^{***}	28.0	0.0001 ^{***}	0.024 [*]
Scrt9	Height	0.055 [†]	20.0	NS	0.051 [†]

†, *, **, *** Significant at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

NS, Not Significant at the 0.10 level.

Table A.6. cont'd. Summarized List of Traits per Marker for F₂ Population at Knoxville, TN.

Marker	Trait	Marker p-value	r ² (%)	Additive p-value	Dominant p-value
Satt50	Maturity	0.0008***	7.3	0.0002***	NS
Satt353	----	----	----	----	----
Satt30	Yield	0.089 [†]	14.1	NS	0.052 [†]
	Height	NS	17.9	0.095 [†]	NS
Satt251	Seed Protein	0.081 [†]	2.7	0.029*	NS
	Seed Oil	0.031*	3.8	0.012*	NS
Satt14	Seed Oil	0.011*	5.3	0.0088**	0.033*

†, *, **, *** Significant at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

NS, Not Significant at the 0.10 level.

Table A.7 Summarized List of Traits per Marker for the F_{4:6} Population at Knoxville, TN.

Marker	Trait	Marker p-value	r ² (%)	Additive p-value	Dominant p-value
Satt540	Lodging	NS	2.1	NS	0.082 [†]
Satt52	Yield	0.085 [†]	3.2	0.044 [*]	NS
	Seed Size	NS	3.8	NS	0.099 [†]
Satt239	Height	NS	7.9	0.045 [*]	NS
Satt373	Seed Protein	0.018 [*]	17.0	0.048 [*]	0.045 [*]
	Height	NS	7.7	NS	0.070 [†]
Sctt9	Seed Oil	NS	6.5	NS	0.048 [*]
Satt50	Seed Protein	0.071 [†]	15.1	NS	0.068 [†]
Satt353	Seed Size	0.034 [*]	5.7	0.069 [†]	0.071 [†]
Satt30	----	----	----	----	----

†, *, **, *** Significant at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

NS, Not Significant at the 0.10 level.

Table A.7 cont'd. Summarized List of Traits per Marker for F_{4:6} Population at Knoxville, TN.

Marker	Trait	Marker p-value	r ² (%)	Additive p-value	Dominant p-value
Satt251	----	----	----	----	----
Satt14	Seed Protein	0.011*	51.0	NS	0.012*
	Seed Size	0.0001***	15.8	0.034*	0.0002***
	Lodging	NS	2.6	0.077†	NS

†, *, **, *** Significant at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

NS, Not Significant at the 0.10 level.

Table A.8 Summarized List of Traits per Marker for the F_{4:6} Population at Highland Rim Experiment Station.

Marker	Trait	Marker p-value	r ² (%)	Additive p-value	Dominant p-value
Satt540	----	----	----	----	----
Satt52	----	----	----	----	----
Satt239	Height	0.088 [†]	3.1	0.030 [*]	NS
Satt373	Height	NS	2.5	NS	0.070 [†]
Sctt9	----	----	----	----	----
Satt50	Seed Size	NS	2.0	0.078 [†]	NS
Satt353	Yield	NS	2.4	0.051 [†]	NS
	Seed Protein	NS	2.4	0.090 [†]	NS
	Seed Oil	NS	2.5	0.067 [†]	NS

†, *, **, *** Significant at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

NS, Not Significant at the 0.10 level.

Table A.8 cont'd. Summarized List of Traits per Marker for F_{4:6} Population at Highland Rim Experiment Station.

Marker	Trait	Marker p-value	r ² (%)	Additive p-value	Dominant p-value
Satt30	----	----	----	----	----
Satt251	Seed Protein	NS	2.4	0.077 [†]	NS
	Seed Oil	NS	2.6	NS	0.078 [†]
Satt14	Seed Size	0.016 [*]	5.5	NS	0.023 [*]

†, *, **, *** Significant at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

NS, Not Significant at the 0.10 level.

Table A.9 Summarized List of Traits per Marker for the F_{4:6} Population Combined over Knoxville and Springfield, TN.

Marker	Trait	Marker p-value	r ² (%)	Additive p-value	Dominant p-value
Satt540	----	----	----	----	----
Satt52	Yield	NS	1.8	0.086 [†]	NS
Satt239	Height	NS	2.6	0.046 [*]	NS
Satt373	Seed Protein	0.042 [*]	3.8	NS	0.028 [*]
	Height	NS	2.4	NS	0.071 [†]
Sc tt9	Seed Oil	NS	2.7	NS	0.053 [†]
Satt50	----	----	----	----	----
Satt353	Seed Size	NS	2.5	0.093 [†]	NS

†, *, **, *** Significant at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

NS, Not Significant at the 0.10 level.

Table A.9 cont'd. Summarized List of Traits per Marker for F_{4:6} Population for Both Locations.

Marker	Trait	Marker p-value	r ² (%)	Additive p-value	Dominant p-value
Satt30	----	----	----	----	----
Satt251	Seed Protein	NS	2.6	0.040*	NS
Satt14	Seed Protein	0.081 [†]	3.4	NS	0.033*
	Seed Size	0.0002***	11.3	0.073 [†]	0.0008***

[†], *, **, *** Significant at the 0.10, 0.05, 0.01, and 0.001 levels, respectively.

NS, Not Significant at the 0.10 level.

Table A.10 Positive and Negative Parental Alleles, Linkage Groups, Base Pairs, and Trait Means for QTL's for an F_{4:6} Essex x Williams Population Combined over Knoxville and Springfield, TN.

Significant Trait	Marker	LG	+ Allele	Base Pairs	Mean	- Allele	Base Pairs	Mean
Yield	Satt52	H	Williams	246	305 g/plot	Essex	241	288 g/plot
Seed Protein	Satt251	B1	Essex	216	44.1 %D W	Williams	226	43.6 %D W
Seed Size	Satt353	H	Essex	174	12.3 g/100 seed	Williams	165	11.8 g/100 seed
	Satt14	D2	Essex	205	12 g/100 seed	Williams	183	11.6 g/100 seed
Height	Satt239	I	Williams	205	27.2 cm	Essex	208	24.9 cm

+ = Positive QTL, denoting significantly higher gelscore class mean

- = Negative QTL, denoting significantly lower gelscore class mean

LG = USDA-ARS, ISU map linkage group

VITA

The author, Angie Bell Chapman, was born in Jackson, Tennessee on August 16, 1972 to Don and Avis Bell. She has one younger sister, Amy. Angie enjoyed riding her horses while growing up in Jackson and spent as much time as possible outdoors.

Angie was married to Brian Paul Chapman on March 24, 1998 in the small town of Moab, Utah. The author and her husband spend their spare time traveling, mountain biking, and hiking with their two dogs, Guyot and Sierra.

