To the Graduate Council:

I am submitting herewith a dissertation written by Ali Ustun entitled "Variations for some quantitative traits and molecular markers in some major ancestors and cultivars of soybean, and evaluation of yield increase by breeding in the southern U.S." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Plant and Soil Science.

Fred L. Allen, Major Professor

We have read this dissertation
and recommend its acceptance:

Accepted for the council:

Associate Vice Chancellor and
Dean of The Graduate School

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.

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Ali Ustun

August 1999

DEDICATION

This thesis is dedicated to my mother
Fatma Ustun who passed away when I just started PhD
studies in the U.S.

And

my son Alptekin Ustun who was born
during my study for this degree

ACKOWLEDGEMENTS

There are many people whom I am grateful for their help and support at the University of Tennessee. However, the very first person, I am very indebted to, is Dr. Fred L. Allen who is the head of Plant and Soil Science Department and my major professor for his consistent consultation, support and encouragement not only in the preparation of this thesis but also in other aspects of the plant breeding, genetics and life. There is little I can write here to express my feelings for him, but he was a friend, a father, a brother and a supervisor based on the situation. It was my privilege to be a student of his.

I wish to thank Dr. Peter M. Gresshoff for his consistent advice and challenges on molecular marker technology. I would like to thank Dr. Arnold M. Saxton for his open door policy to help and teach me statistics. May I thank my other two committee members, Dr. Burton C. English and Dr. Dennis W. West for their continuous help and their patience with me. I should also acknowledge help and friendship of Dr. Vincent Pantalone.

There are a number of people whose assistance should be recognized. I would like to thank Mrs Debbie Landau-Ellis, Richard Johnson, Carolyn Meyer and all personnel of Knoxville Experiment Station for their help in field work and friendship.

I should not forget to thank the staff and graduate students in Plant Molecular Genetics Laboratory for their moral support, friendship and endless discussions on scientific matters.

I would like to express my gratitude to the faculty, staff and students of the Department of Plant and Soil Sciences for making my experience there one that I will cherish.

The last but not the least, the greatest debt is owed to my wife, Meral, my daughter, Damlanur and my son, Alptekin for their patience and missing me at home without unfailing support and encouragement. Without them, I would not be happy during my stay in Knoxville.

Abstract

Formal soybean breeding programs in the U.S. started in the 1930's and extensive research changed the plant from a forage crop to a seed crop. Soybean has become one of the five crops commercially important. Soybean breeders have made selection mostly for yield, protein and oil content, lodging, and disease and pest resistance. Genetic progress in these traits might have caused decline in genetic variability. On the other hand, different soybean breeding lines have been crossed in the southern and northern U.S. This study was aimed to (1) estimate the yield increase in the southern U.S. by genetic improvement, (2) show genetic variability of agronomic traits and find molecular marker diversity in soybeans from ancestral lines to modern cultivars, and (3) demonstrate the genetic variability by heritabilities and Quantitative Trait Loci (QTL) in eight crosses made between southern x southern and northern x southern ancestral lines or cultivars representing different eras in soybean breeding.

A total of 10 yield trials were conducted in two years in four locations in Tennessee with eight ancestral lines and cultivars from different eras. Soybean yield in the southern U.S. has increased 14 kg $ha^{-1}yr^{-1}$.This increase was linear which implies that there is still remaining variation in the southern soybean gene pool to make genetic progress for yield.

Genetic diversity among eight southern and four northern cultivars was determined by microsatellite (SSR) and DNA Amplification Fingerprinting (DAF) molecular markers. Also, data from eight F_2 populations and parents were collected. SSR and DAF marker polymorphisms, and heritabilities for yield and oil content showed a declining trend over generations in the southern breeding lines. Correlations among genetic similarities from SSR, DAF, CP (Coefficient of parentage) and agronomic traits were significant. Although SSR similarity was correlated with yield, it should be retested with a larger sample size. It is concluded that the markers which have QTL associations with agronomic traits may provide very useful information to decide parents for crosses.

$F_{2:3}$ lines derived from individual 250 F_2 plants in each cross were tested in the field

for agronomic traits. With the exception of Lee x Ogden (first generation cultivars) cross, other southern x southern crosses did not reveal a significant heritability for yield from parent-offspring regression. Northern x southern crosses had considerably higher heritability for yield, but their population mean was lower than southern x southern crosses. It is suggested that the elite lines from Lee x Ogden and from recurrent selections in northern x southern crosses can increase variability for yield in the southern gene pool.

One cross of southern ancestral lines and one cross of fifth generation cultivars were also evaluated for SSR marker QTL associations with agronomic traits. In the cross between ancestral lines, SSR markers Satt022, Satt180 and Satt353 were associated with yield in F_2 but Satt022 was not confirmed in $F_{2:3}$ lines. There were QTLs for other traits which were either significant or near to significant level either in F_2 or in F_3. In the cross between fifth generation cultivars, Satt022 was associated with yield and Satt353 was associated with oil content and plant height. The change in the QTL relationships from ancestral lines to the fifth generation cultivars implies that some QTLs may not be detected, but different QTLs were developed over time in southern breeding lines. The variability for yield, which was not detected by heritabilities as found by SSR marker QTLs. This indicates that QTLs can be a reliable way to demonstrate genetic variability.

PREFACE

Soybean is a very good example for the achievement of research in our daily life. It was a forage crop before the 1930's. When its oil and protein values were realized, intensive breeding and crop management research converted this forage crop to first an important food and feed seed crop. From oil and meal products, it has a very wide usage, maybe the widest usage among crops. Soybean breeders in the U.S. started to increase the yield and quality components of crop. In addition to their efforts, agronomists, pathologists, nematologists and other related scientists contributed to the yield increase and expanding production. As a result, soybean yields have doubled in the U.S. in the last 50-60 years.

Soybean cultivars in the U.S. have been developed from two distinct gene pools. Maturity groups 00-III have been developed from a "northern" gene pool while maturity groups V-VIII have been developed from a "southern" gene pool. Cultivars of maturity group IV have been developed from both gene pools. While germplasm collections available for use in those gene pools are large, only a smaller number of ancestral lines were utilized in the beginning to establish the "gene pool" for northern and southern soybean breeding programs in the U.S. After the first generation cultivars were developed, most breeding efforts have focused on crossing elite by elite lines tracing to the original narrow germplasm bases. The combined number of ancestral lines for the northern and southern gene pools is approximately 50 with 14 of those contributing about 80 % of the genes to the current gene pool. The focus of this study is to evaluate the contribution of plant breeding to the genetic progress that has been made in breeding for yield in southern cultivars (particularly maturity groups V and VI), and to evaluate genetic variation remaining for yield and other important agronomic traits given the narrow genetic base. The approach is the recreate a cross between the two major southern ancestral lines and then compare the genetic variation and performance with crosses between selected first, third and fifth generation cultivars that have been derived from them. These results were then compared with a cross between a major northern ancestral line as well

as first, third and fifth generation southern cultivars crossed with first, third and fifth generation northern cultivars. A major consideration here is what genetic variation might have been possible from crosses between major cultivars derived from the separate gene pools.

In all experiments, the ancestral lines and cultivars were unique because: (1) they represented different eras of production from 1940's to 1950's, (2) ancestral lines are major contributing lines to the modern gene pool of current cultivars, (3) the cultivars represent different generations of related cultivars and they were widely grown during their era of production, and (4) most of them have been utilized extensively as parents in both public and private breeding programs.

Part I at this dissertation contains the experiments which evaluated genetic progress for southern soybeans. Similar studies have been conducted for northern soybean cultivars and illustrated the annual gain from genetic improvement of soybean. In the south, a similar study was conducted, but it covered cultivars from maturity group (MG) VI to VIII. There has not been any investigation for MG V-VI soybean which is grown in the mid-southern states.

Part II shows the results of estimating genetic variance via DNA markers such as microsatellite (SSR) and DNA Amplification Fingerprinting (DAF) and agronomic trait variation in F_2 generation. Crosses were made between the two southern ancestral lines and between one southern and one northern ancestral lines. Crosses were also made between first, third and fifth generation southern x southern or northern x southern cultivars. This is the first example of evaluating sequential crosses between major cultivars representing different decades of soybean breeding. Soybean genetic variability from DNA markers and quantitative trait variation were evaluated from ancestral lines to the fifth generation cultivars. DNA markers have been applied to plant breeding research for marker assisted selection (MAS), genetic mapping and genotype fingerprinting. However, they were not used to illustrate how soybean genetic variability changed over time. This study applied DNA markers as a tool to approach genetic variability in soybean history in the southern U.S. SSR and DAF marker

data also showed that coefficient of parentage assumptions (zero coefficient between "unrelated" genotypes in the pedigree and equal contribution of parents to the progeny) are not true and can be adjusted by the information from marker data. In addition to these, choice of parents and cultivar classification by cluster analysis were also discussed.

Quantitative Trait Loci (QTL) which is the association of a genetic marker with a quantitative trait has application to locate and map genes for complex traits like yield. However, they were not used to address genetic variability in crops. Part III deals with QTLs related to genetic variability changes in soybean. Besides QTL, narrow sense heritabilities were estimated in eight populations, which were also materials for Part II, in F_2 and F_3 populations. QTL data from a cross between the two southern ancestral lines and a cross between fifth generation cultivars were compared to illustrate changes over time.

TABLE OF CONTENTS

CHAPTER PAGE

PART I: Genetic Progress in Southern Soybean Cultivars

PART II: Assessment of Genetic Diversity Among Soybean Cultivars from Different Eras by SSR and DAF Markers and by Agronomic Traits

PART III: Genetic variation in F_2 and F_3 Soybean Populations in North America by QTL Associations with Agronomic Traits

PART IV: Summary and Future Research

LIST OF TABLES

TABLE PAGE

PART I: Genetic Progress in Southern Soybean Cultivars

PART II: Assessment of Genetic Diversity Among Soybean Cultivars from Different Eras by SSR and DAF Markers and by Agronomic Traits

PART III: Genetic variation in F_2 and F_3 Soybean Populations in North America by QTL Associations with Agronomic Traits

LIST OF FIGURES

FIGURE PAGE

PART I: Genetic Progress in Southern Soybean Cultivars

PART II: Assessment of Genetic Diversity Among Soybean Cultivars from Different Eras by SSR and DAF Markers and by Agronomic Traits

PART III: Genetic variation in F_2 and F_3 Soybean Populations in North America by QTL Associations with Agronomic Traits

PART I

GENETIC PROGRESS IN SOUTHERN SOYBEAN CULTIVARS

Abstract

Yield increase in crops has occurred due to plant breeding, production and management techniques. Eight cultivars representing different eras of soybean breeding in southern and mid southern states from 1940s to 1990s were tested in four locations for three years in Tennessee in order to evaluate the genetic improvements in yield and other important agronomic traits. CNS and S.100 were chosen as representative ancestral lines from the 1940s. Ogden and Lee, Hill and Essex, and TN 5-85 and Hutcheson were released in 1950s, late 1960s and early 1970s, and late 1980s, respectively. The experiments were conducted in Knoxville and Ames for three years, and Milan and Springfield for two years. The results showed that genetic improvement shortened the plant height and decreased lodging. Oil content was increased until the 1980s, but then started to decrease; however, due to correlated effects, protein content started to increase after the 1980s. Yield increase over time was linear and based on genetic variation present within the southern soybean gene pool, additional gain from selection for yield can be achieved. Crop improvement increased soybean yield 14 kg ha^{-1} per year. Current cultivars showed better yield stability and more response to better growing conditions than ancestral lines.

CHAPTER 1

Introduction and Literature Review

One way to increase yield significantly is through genetic manipulation of crops. Soybean breeders have been able to increase the seed yield and improve other traits like lodging, protein content and oil content. Soybean yield increased from 1350 kg ha^{-1} to 2250 kg ha^{-1} in the last 60 years in the U.S. (FAO production yearbooks 1948 to 1994). This considerable increase cannot be attributed to only cultivar improvement, because management practices also played a significant role in yield increase.

Accurate estimation of genetic progress realized over time for a given trait is a difficult task. The number of genotypes and choice of cultivars for the experiment may change the estimate. Cox *et al* (1988) suggested that evaluation of cultivars in common environments provided the most direct estimate of breeding progress. Slafer *et al* (1994) stated that an experiment for this purpose should satisfy the following conditions: (1) experiments must be conducted under field conditions, (2) measurements must be made on plots, not on single plants, and (3) cultivars released at different times must be compared simultaneously.

In the north central region of the U.S., the contribution of soybean breeding to yield increase was between 10 and 21 kg ha^{-1} per year from 1902 to 1977 in maturity groups 00 through IV (Specht and Williams, 1984). However, when they summarized only cultivars from hybridizations after 1939, the average annual gain was 12 kg ha^{-1}. Williams and Specht (1979) concluded based on the same study that 79% of the yield increase in the northern states could be attributed to genetic improvement. Wilcox *et al* (1979) compared five cultivars from both maturity groups II and III. They found that the cultivars gave 25% more seed yield compared to ancestral lines. Protein content decreased over time, but this was accompanied by an increase in oil content. They found similar regression coefficients for cultivars in estimation

of yield stability.

Luedders (1977) tested 21 first and second cycle soybean cultivars to estimate genetic improvement. It was estimated that increase in yield for the first and second cycles was 26 and 16%, respectively. He showed that lodging resistance was improved substantially in commercial cultivars in maturity groups II, III and IV. During this time, only a slight increase in plant height occurred.

Boerma (1979) tested 18 southern cultivars from maturity groups VI, VII and VIII. He reported significant yield increase over time and from 1942 to 1973, this increase was 0.7% (13.7 kg ha^{-1}) per year. This yield increase was associated with increasing pod number in soybeans. Seed size and number of seeds per pod did not show significant change over time. Plant height was decreased in maturity group VIII but did not change in groups VI and VII.

Karmakar and Bahatnagar (1996) compared cultivars developed from 1969 to 1993 in India. They found that the annual genetic gain was 22 kg ha^{-1}. Voldeng *et al.* (1997) tested 41 soybean cultivars from maturity group 0, 00 and 000 to show the genetic improvement in soybeans in Canada. These cultivars were released between 1934 and 1992. They found that annual yield increase was 0.5 % (9.3 kg $ha^{-1}yr^{-1}$) until 1976 and after that it was 0.7 % (13 kg $ha^{-1}yr^{-1}$). In their study, it was observed that lodging tolerance was increased without any decrease in plant height. Yield stability of cultivars was constant over time.

Similar studies have been conducted in other crops. Waddington *et al.* (1986) conducted experiments on bread wheat with cultivars from 1950 to 1982. They found that the yield gain per year was 59 kg ha^{-1} in Mexico. They also observed that yield potential had plateaued in the early 1970s, but yield improvement returned the same rate after 1979 when genetic variability was broadened. Schmidt (1984) and Schmidt and Worrall (1984) indicated a possible yield plateau in wheat in the late 1970s. Cox *et al.* (1988) showed that genetic improvement of yield was restored by an increase in genetic diversity in wheat in 1980s. They found that wheat yield increased 16.2 kg ha^{-1} per year from 1874 to 1987. They included 35

hard winter wheat cultivars in their study. Lodging resistance is another trait which has improved in wheat over time. Sayre *et al.* (1998) compared 15 wheat cultivars released between 1966 and 1988 in Mexico. The yield increase was 0.48% and yield progress was more dramatic when slow rusting resistance genes were incorporated to cultivars.

In the mid-southern states of the U.S., the soybean acreage has decreased since the 1970s, although yield improvements have occurred in soybean. It is not known how much of the yield increase has been caused by genetic improvement or advances in technology or growing soybean in better areas. This study was conducted to estimate (1) the contribution of public soybean breeding programs to yield increase, and (2) what changes occurred in other important agronomic traits.

CHAPTER 2

Materials and Methods

Two ancestral lines and six soybean cultivars representing different eras of development were chosen for this study. Ancestral lines, and first, third and fifth-generation cultivars were represented by two cultivars which have been produced on significant acreage in southern states (Table 1).

Table 1. Soybean ancestral/lines cultivars, their release years, cultivar generations and maturity groups.

Cultivar	Release Year	Generation	Maturity
CNS	Introduction	Ancestral Line	VII
S.100	Introduction	Ancestral Line	V
Lee	1954	First	VI
Ogden	1953	First	VI
Hill	1959	Third	V
Essex	1972	Third	V
TN 5-85	1986	Fiffth	V
Hutcheson	1987	Fitfh	V

The experiments were conducted in a randomized complete block design with three replications. Experiments were carried out in Knoxville and Ames Plantation (1996-1998), Milan and Springfield (1997 and 1998). Cultivars were seeded in four rows 6.1 m long with a 76 cm spacing between rows. Experiments were usually established in mid-May with the

exception of Milan where was seeded in mid-June in 1998.

Maturity, plant height, lodging, protein and oil content, and yield were recorded. Maturity was recorded when 95% of the plants were ready for harvest and it was the number of days between planting and physiological maturity. Plant height was measured at maturity as being the distance between soil surface and the apex of the main stem. Protein and oil were analyzed with a Model 1255 Infratec NIR food and feed analyzer at National Center for Agricultural Utilization Research, Peoria, Illinois. Protein and oil contents were adjusted for dry weight. Lodging was scored on 1 to 5 scale; 1 being all plants erect and 5 being 95+% of plants prostrate. Yield data were taken from the two center rows of each plots after 50 cm from the ends of the rows were discarded. Yields were adjusted to 130 g kg^{-1} moisture content.

SAS Procedures MIXED, GLM and REG were run to obtain analysis of variance for different traits and to estimate regression parameters, respectively. In the mixed model, cultivars and locations were designated as fixed factors and years and blocks as random factors. In combined analysis of variance, each experiment was treated as one environment. In regressions for the genetic trends, least square mean of agronomic traits were used as dependent variable. Stability analyses were carried out by taking experiment means as environmental index (Eberhart and Russell, 1966). The mean yield of each cultivar in each experiment was regressed on the environmental index.

CHAPTER 3

Results and Discussion

Cultivars differed from each other significantly for all traits. Significant differences were also found between environments which was the pooled effects of years and locations. Cultivar by environment interactions were significant which shows the magnitude of genotype by environment interaction(Table 2).

Table 2. Mean squares (MS) and degrees of freedom (df) for the combined analysis of variance of yield, protein content and oil content for soybean cultivars/lines representing different eras in the mid-southern states of the U.S.

	Protein content		Oil content		Yield	
Source	df	MS	df	MS	df	MS
Environment (E)	6	128.3**	6	35.8**	9	16829556.0**
Blocks/E	14	1.4	14	0.3	20	141220.2
Cultivars (C)	7	26.4**	7	22.1**	7	2791103.4**
C * E	40	2.7**	40	1.2**	62	373399.8**
Error	90	0.7	90	0.3	134	116081.6
CV (%)		2.0		3.1		15.0
R^2		0.94		0.94		0.93

** Significantly different from each other ($P \leq 0.01$)

Yield has been the major objective in most breeding programs. This study demonstrates that significant yield increases were achieved by soybean breeders in the southern U.S. The mean yield of TN 5-85 and Hutcheson (Fifth generation cultivars) was

Table 3. Some agronomic traits of soybean cultivars and ancestral lines representing different eras in the mid-southern states of the U.S

Cultivar	Maturity** (Days)	Height** (cm)	Lodging** (1-5)	Protein** (%)	Oil** (%)	Yield** (kg ha^{-1})
CNS	153	87.8	3.6	44.5	16.2	1968.5
S.100	125	105.7	3.0	45.6	17.4	1974.7
Lee	143	81.8	2.8	42.8	19.7	2197.1
Ogden	140	89.6	2.4	43.2	18.8	2151.5
Hill	126	74.5	2.7	42.4	18.9	1952.7
Essex	127	67.1	1.4	44.2	18.8	2520.9
TN 5-85	131	83.0	2.0	43.5	18.8	2487.0
Hutcheson	132	77.1	1.6	42.1	19.8	2731.4

** Significantly different from each other ($P \leq 0.01$)

32.3 %, 20.0 % and 16.6 % higher than that of ancestral lines, and first and third generation cultivars (Table 3). Soybean yield was increased from 1971.6 kg ha^{-1} to 2609.2 kg ha^{-1}. The dramatic increase occurred with the improvement of Essex. Essex in the southern states and Williams and Amsoy in the northern states could be addressed as cornerstones in the soybean breeding history. Remarkably, these three cultivars were released in the same decade (1970s).

The annual gain was 14 kg ha^{-1} and the increase is still maintained (Figure 1). However, cultivars accounted for only 7.0 % of the yield variation. It seems that it was due to large environmental effect on the yield. Full model explained 93 % of the variation on yield (Table 2). The rate of gain in this study is lower than that reported by Karmakar and Bahatnagar (1996), and it is consistent with the rate of Luedders (1977), Wilcox *et al.* (1979) Boerma (1979), and Specht and Williams (1984). In general the annual genetic gain for yield in the southern and northern U.S. is similar. The rate of progress for yield in soybeans is lower

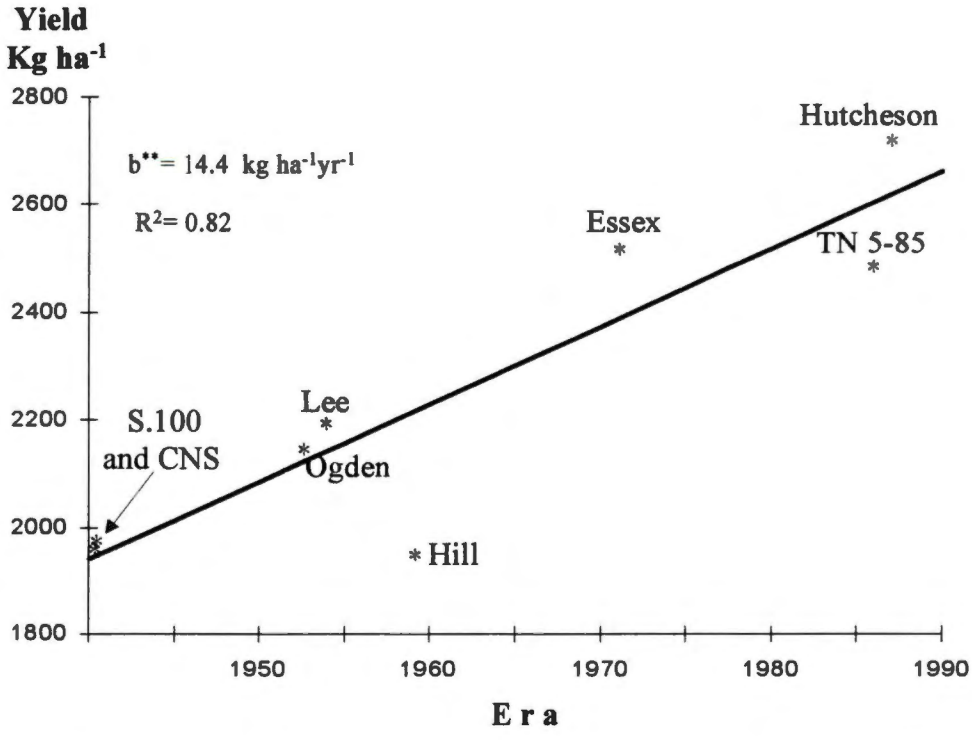


Figure 1. Yield increase of soybean ancestral lines/cultivars representing different eras in the mid-southern states of the U.S.
** significantly different from zero ($P \leq 0.01$)

than that in wheat. Waddington *et al.* (1986) showed a 59 kg ha^{-1} yield increase per year in wheat. They included cultivars developed after the green revolution.

After the corn blight epidemic in 1972, narrow genetic variability became a concern. Furthermore, sometimes it has been claimed that genetic improvement has reached the yield plateau, as in wheat (Waddington *et al.* 1986, Schmidt, 1984 and Schmidt and Worrall, 1984). Neither this study nor other previous studies have shown any indication for a yield plateau in soybeans. It does not mean that broadening the genetic variability in soybeans should be neglected. The linear increase of yield over time implies that if similar breeding strategies were applied, it would be expected that yield increase would be maintained. In other words, it might be stated that considerable genetic variation remains among southern breeding lines to make genetic progress in soybeans. Another implication of this result is the

importance of conventional plant breeding. New biotechnological tools such as marker assisted breeding and transgenic cultivars may accelerate the rate of genetic progress. When the need to increase world food supply, the results of this study and those of previous studies were considered, it is concluded that plant breeding programs should be kept active.

Protein and oil content are both very important traits, but they are negatively correlated and negatively correlated with yield. This correlation can be seen in Figure 2 where traits show a reversed trend. It seems that emphasis was to increase oil content at the expense of protein until 1980s, when oil content reached around 19.5 % in the southern cultivars and after that emphasis was placed on increasing protein content. However, the oil content of cultivars showed a different picture. Hutcheson had the highest oil content, and Ogden, Hill, Essex and TN5-85 yielded almost the same percent oil content (Table 3). The largest difference was between ancestral lines and cultivars. Wilcox *et al.*(1979) showed that oil content in soybeans was increased, but protein content was decreased. Considering that their study was conducted before 1980, our results and their results are consistent for the same period. In these two traits negative correlations and low genetic variation for selection seem to limit the genetic progress. Although the negative correlation exists between two traits, it is possible to create a cultivar with relatively high protein and oil content such as Essex.

Over generations, maturity shifted towards earlier genotypes. As a result, days to maturity declined significantly. From ancestral lines to the fifth generation cultivars, maturity was shortened by 10 days (Table 3). Early cultivars have less harvest risks compared to late cultivars due to leave the field early and have less moisture content, and also they also give producers a flexibility to adjust labor. Decrease in maturity did not happen on the expense of seed yield. There is no other research result which supports or conflicts with this finding.

Plant height and lodging are interrelated in most cases. In southern U.S., plant height was decreased until 1970's, but it began to increase after that (Figure 3). This might be due

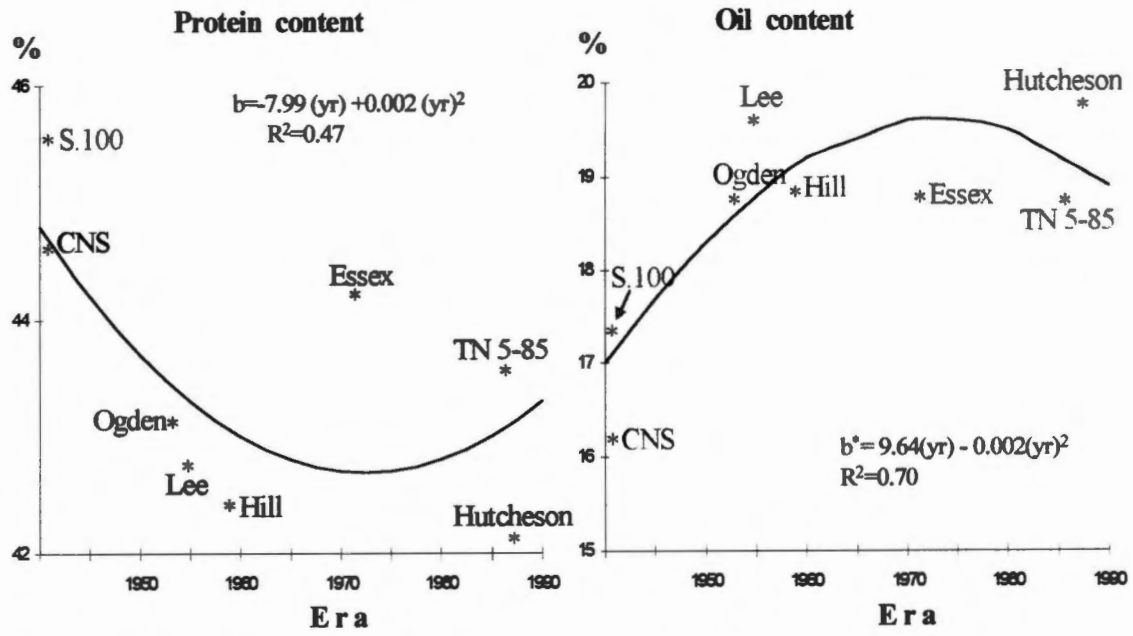


Figure 2. Changes in protein and oil content of soybean cultivars/ancestral lines representing different eras in mid-southern states of the U.S.

*linear and quadratic regressions significantly different from zero ($P \leq 0.05$)

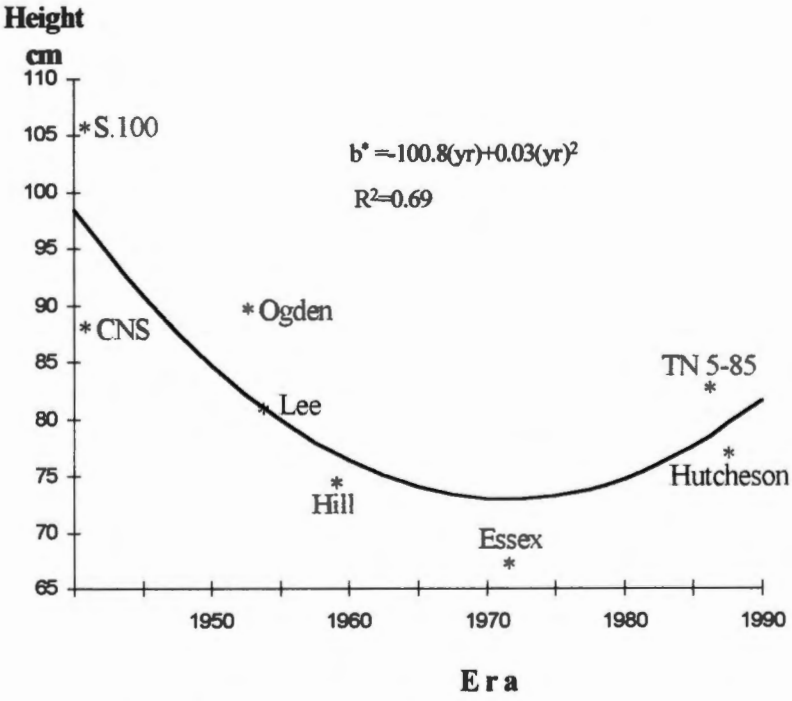


Figure 3. Changes in plant height of soybean cultivars/ancestral lines representing different eras in mid-southern states of the U.S.

*linear and quadratic regressions significantly different from zero ($P \leq 0.05$)

to increase in lodging resistance in taller plants. When sufficient success has been achieved for lodging, a good strategy would be to increase plant height, leading to higher yield potential. Soybean breeders first selected shorter plants to reduce lodging. Plant height was approximately 97 cm in ancestral lines, and it was around 80 cm in the fifth generation cultivars. Essex had the shortest plants (Table 3). Plant height was slightly increased in cultivars tested by Luedders (1977) and Voldeng *et al.* (1997). Their material was consisted of maturity group 000 to IV. Within the same environment, early cultivars tend to have shorter plants; therefore, plant height may not be associated with lodging in early cultivars. In northern material, less emphasis might have been given to plant height compared to southern soybean material. As a matter of fact, Boerma (1979) showed that only in maturity group VIII, plant height was decreased. Lodging showed a significant linear decrease over time. Lodging was a trait in all research that was decreased in soybeans (Ludders, 1977, Voldeng *et al.* 1997) and in wheat (Cox *et al.* 1988). Both plant height and lodging are the major plant traits which affect the type of harvest and harvest losses. Data from this research and other studies imply that breeders spent some effort to increase yield by avoiding losses and by developing plants which are suitable for mechanical practices.

Stability of Performance

When yield stability of cultivars in this experiment were examined in terms of stability coefficients (regression coefficient), the apparent difference was between cultivars and ancestral lines; ancestral lines having a coefficient less than one, and cultivars having coefficients greater than one, with the exception of Hill. Wilcox *et al* (1979) and Voldeng *et al.* (1997) obtained results which conflict with our findings for yield stability. The reason for this conflict might be different cultivars and test locations in those studies. A coefficient less than one reflects that genotype is less responsive as the environment changes from low yielding to high yielding. Ancestral lines came out as a result of natural selection; therefore, they might be less responsive to changes in environment compared to cultivars. Frederick *et*

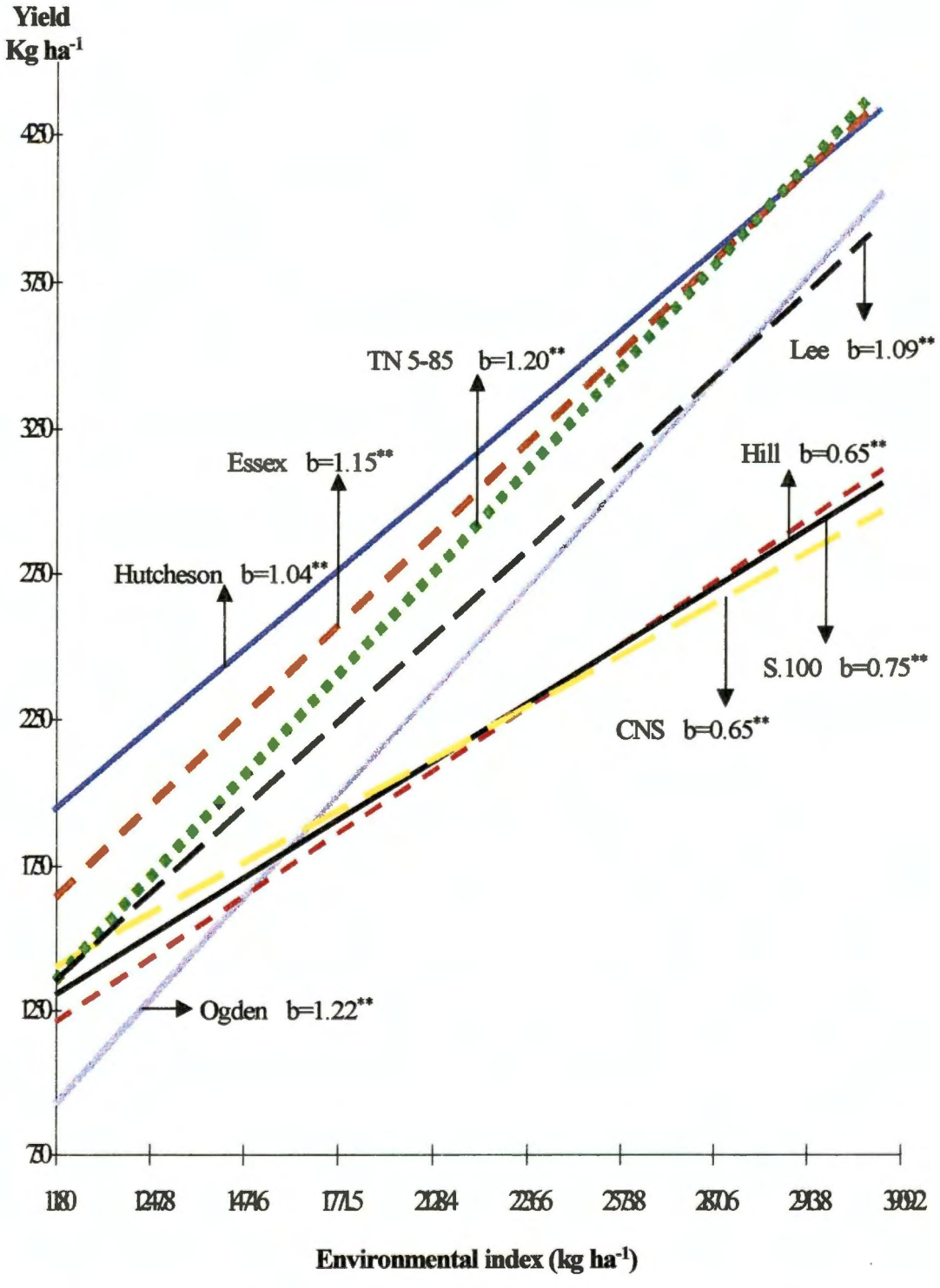


Figure 4. Yield stability of soybean cultivars representing different eras in the southern and mid-southern states

** significantly different from zero ($P \leq 0.01$)

al. (1990 and 1991) compared two old (Manchu and Dunfield) and two modern (Williams 82 and Clark 63) soybean cultivars under drought stress and irrigated conditions. Although older cultivars were found more efficient in conserving water under drought, the yield increase from old cultivars to modern cultivars was 31 and 9 % under irrigated and drought conditions, respectively.

Most of the cultivars in this experiment had almost the same yield level in low yielding environments, but Hutcheson and Essex gave the highest yield and Ogden gave the lowest in these conditions. In higher yielding environments, yield differences were more pronounced. Figure 4 shows that among these eight cultivars, Hutcheson was the most desirable cultivar for yield and yield stability, because it had higher or equal yield in any given environment than the others. In higher yielding environments, latest cultivars (Hutcheson, TN 5-85 and Essex) were superior to older cultivars. Either Essex or TN 5-85 are comparable to Hutcheson in higher yielding environment.

CHAPTER 4

Conclusions

Genetic improvement of soybeans in the southern states has led to improved plant architecture; therefore, dense planting became possible due to tolerance to lodging. Improvement of early maturity type cultivars have supplied a kind of natural insurance against the risk of late harvest. It is difficult to assess the direct and accurate contribution of genetic improvement to yield, but average yield of current cultivars tested was approximately 27 % higher than ancestral lines and 20 % higher than first generation cultivars. Breeding for oil and protein content at the same time represents a paradox. This dilemma between two negatively correlated traits has been shown in this research. For specific needs, high protein or high oil cultivars can be developed at the expense of the other. However, breeders have succeeded at least to maintain the protein and oil content within desirable limits when they have increased yield. Compared to ancestral lines, current cultivars have increased potential to produce higher yield in response to improved growing conditions.

The linear increase in yield over time emphasizes the steady genetic progress that soybean breeders have made in developing cultivars for the mid-south. This linear increase should not be taken as evidence of no need for increased variation. At the moment, crossing within southern lines probably will give satisfactory gain from the selection. However, continuous selection on the same material for long period may cause very narrow genetic variability which can induce yield plateau. The strategy for soybean breeding should consider two objectives at the same time. The first one is to use available adapted elite germplasm to develop cultivars as in the past, and the second one is to broaden the genetic base within southern gene pool by making crosses with elite northern lines and plant introductions.

LITERATURE CITED

Boerma, H. R. 1979. Comparison of past and recently developed soybean cultivars in maturity groups VI, VII and VIII. Crop Sci. 19 : 611-613

Cox, T. S., J. P. Shroyer, Liu Ben-Hui, R. G. Sears and T. J. Martin. 1988. Genetic improvement in agronomic traits of hard red winter wheat cultivars from 1919 to 1987. Crop Sci. 28 : 756-760

Eberhart, S. A. and W. A. Russell. 1966. Stability parameters for comparing varieties. Crop Sci. 6 : 36-40

Frederick, J. R., J. T. Woolley, J. D. Hesketh and D. B. Peters. 1990. Seed yield and agronomic traits of old and modern soybean cultivars under irrigation and soil water-deficit. Field Crops Res. 27 : 71-82

Frederick, J. R., J. T. Woolley, J. D. Hesketh and D. B. Peters. 1991. Water deficit development in old and new soybean cultivars. Agronomy Journal 82 : 76-81

Karmakar, P. G. and P. S. Bhatnagar. 1996. Genetic improvement of soybean varieties released in India from 1969 to 1993. Euphytica 90 : 95-103

Luedders, V. D. 1977. Genetic improvement in yield of soybeans. Crop Sci. 17 : 971-972

Sayre, K. D., R. P. Singh, J. Huerta-Espino and S. Rajaram. 1998. Genetic progress in reducing losses to leaf rust in CIMMYT-derived Mexican spring wheat cultivars. Crop Sci. 38 : 654-659

Schmidt, J. W. 1984. Genetic contributions to yield gains in wheat. *In* W. R. Fehr (ed.) Genetic contributions to yield gains of five major crop plants. CSSA Spec. Publ. 7. CSSA, Madison, WI

Schmidt, J. W. and W. D. Worrall. 1984. Trends in yield improvement through genetic gains. Proc. 6th Int. Wheat Genet. Symp. 6 : 691-700

Slafer, G. A., E. H. Satorre and F. H. Andrade. 1994. Increase in grain yield in bread wheat from breeding and associated physiological changes. *In* G. A. Slafer (ed.) Genetic improvement of field crops. Marcel Dekker, Inc. pp 1-68

Specht, J. E., and J. H. Williams. 1984. Contribution of genetic technology to soybean productivity-retrospect and prospect. *In* W. R. Fehr (ed.) Genetic contributions to yield gains of five major crop plants. CSSA Publ. 7. CSSA, Madison, WI. p.49-74

Voldeng, H. D., E. R. Cober, D. J. Hume, C. Gillard, and M. J. Morrison. 1997. Fifty eight years of genetic improvement of short-season soybean cultivars in Canada. Crop Sci. 37 : 428-431

Waddington, S. R., J. K. Ransom, M. Osmanzai and A. Saunders. 1986. Improvement in the yield potential of bread wheat adapted to Northwest Mexico. Crop Sci. 26 : 698-703

Wilcox, J. R., W. T. Schapaugh, Jr., R. L. Cooper, W. R. Fehr and M. H. Niehaus. 1979. Genetic improvement of soybeans in the Midwest. Crop Sci. 19 : 803-805

Williams, J. H. and J. E. Specht. 1979. A perspective on yield advances attributable to soybean variety development. Agronomy Abstracts p.81

PART II

ASSESSMENT OF GENETIC DIVERSITY AMONG SOYBEAN CULTIVARS FROM DIFFERENT ERAS BY SSR AND DAF MARKERS, AND BY AGRONOMIC TRAITS

Abstract

Genetic variability and similarity were estimated by agronomic traits, coefficient of parentage and morphological markers until molecular markers became available. Molecular markers can be used to assess genetic trends over time in soybean and other crops. This study was conducted to (1) show genetic trends over time in soybean via microsatellite (SSR) and DNA amplification fingerprinting (DAF) markers, and agronomic traits, (2) evaluate associations between agronomic trait variance and four different GS (Genetic Similarity) estimates, namely SSR-GS, DAF-GS, Coefficient of Parentage (CP) and Quantitative Trait-GS (QT-GS) in eight F_2 populations, and (3) compare GS estimates for cultivar discrimination.

Eight F_2 populations were developed by crossing ancestral lines and cultivars representing different eras. Two sets of four crosses were made between southern and southern, and between northern x southern cultivars/lines. F_2 plants were grown at the Knoxville Experiment Station, and approximately 250 plants from each population and 50 plants from each parent were selected randomly. Maturity, plant height, protein and oil content, and yield were recorded on selected plants. Twelve parents, eight southern and four northern, were assayed by 15 SSR and 15 DAF primers.

Broad sense heritabilities of agronomic traits, and SSR and DAF molecular marker diversity showed that the decline in genetic variability started after the third generation cultivars. There was a considerable amount of variation in ancestral lines. Southern soybean breeding material had low variation for protein and oil content. It seems that there is still variation for yield still remaining to allow genetic progress in the future. Although some significant correlations were obtained between heritabilities and GS estimates, a conclusion could not be drawn due to small sample size. Any study related to marker-agronomic trait variation should be done with markers which show a QTL effect, rather than random markers. SSR and DAF gave better discrimination of genotypes than CP and QT-GS. SSR markers were more efficient than DAF markers to discriminate cultivars. However, SSR markers tend

to underestimate GS. In general, northern and southern soybeans were placed in two different clusters by SSR, DAF, CP and agronomic traits. Some conflicts between clusters, and DAF-GS and SSR-GS were observed. Therefore, research on cluster analysis which was suitable for this type of data is needed. All four GS estimates were correlated very significantly. It was concluded that when pedigree data are available, CP can be used, but CP assumptions of no relationship among ancestral lines and equal contributions of each parent to the offspring can be corrected partly with the information gained from molecular markers.

CHAPTER 1

Introduction and Literature Review

Soybean breeding became very important in the U.S. after the1930s when oil and protein from soybean seed became important products. Up to that time, soybean was mostly used for forage purposes. Researchers, producers and processors, realizing the value of soybean protein and oil content began to promote its production and usage. Soybean breeding programs were initiated to improve the seed yield and seed quality. As a result, soybean yield has increased from 1350 kg/ha in 1948 to 2250 kg/ha in 1994 (FAO production yearbooks from 1948 to 1994). A considerable part of this increase has been due to genetic improvement.

Soybean cultivar improvement activities followed two different pathways in the southern and northern parts of the U.S., although some breeding lines were used in making crosses in both regions. As in most other crops, a limited number of parents was used in crossing programs. Gizlice *et al.* (1993a) reported that only 14 soybean ancestral lines have constituted 70 % to the pedigrees in North America. The southern genetic base was dominated by 17 ancestors, among these CNS and S.100 contributed more than 50 % to the soybean cultivars. Furthermore, a small number of lines have been crossed very frequently in soybean breeding programs in the U.S. (St. Martin, 1982, Delaney *et al.* 1983, Allen and Bhardwaj, 1987, and Gizlice *et al.* 1993a). Soybean breeding became a successful example of recurrent selection in self pollinated crops. Ultimately, a slow but significant yield increase has occurred, but with a decrease in the genetic base. (St. Martin, 1982, Burton, 1987, and Gizlice *et al.* 1994).

Gizlice *et al.* (1994) measured the decline in genetic diversity for soybeans as 21 and 26 % for northern and southern U.S. They found that CNS and S.100 contributed to southern cultivars by 25 % and 21 %, respectively. Contributions of Lincoln, Mandarin (Ottawa) and

Richland to northern cultivars were 24 %, 17 % and 11 %, respectively.

Delaney *et al.* (1983) investigated the relationships among northern ancestral lines and cultivars and found that 12 lines or cultivars contributed 88 % of northern nuclear gene pool while five ancestors accounted for the cytoplasm of 121 of 136 cultivars evaluated. They examined the pedigrees of cultivars from different periods, and showed major contributors of North American soybean cultivars to be Tokyo, PI 54610, CNS and S.100 for southern cultivars, and Mandarin (Ottawa), Mukden, Manchu, Mandarin (Illinois) and Richland for northern cultivars.

Quantification of genetic variability among cultivars or lines and its usefulness in the planning of crosses are important concerns for breeders. So far, the most commonly used method has been pedigree analysis, also known as coefficient of parentage (CP) or coefficient of ancestry. Coefficient of parentage is the probability of an individual getting an allele randomly in a locus, which is identical in descent to one of the parents. Simply it is half the genetic covariance (Allen and Bhardwaj, 1987, Carter *et al.* 1993, and Falconer and Mackay, 1995). However, this estimation of genetic similarity (GS) assumes the absence of selection and genetic drift, which always occur in breeding programs, and zero coefficient for non-related individuals in pedigree. Although CP assumptions cannot be met, and it is a biased estimation in reality, it is still very useful when there are no other available data or method. If the pedigrees of cultivars are known, CP estimation is almost cost free.

Allen and Bhardwaj (1987), and Cater *et al.* (1993) estimated coefficient parentage for 192 and 258 soybean cultivars, respectively. The latter also estimated genetic similarities (GS) between cultivars based on agronomic data. Both studies confirmed frequent use of some lines in breeding programs .

Helms *et al.* (1995) developed six soybean populations by crossing closely related and unrelated parents. They concluded that CP was not useful to predict genetic variance for yield.

Another method, used rarely, is the estimation of similarities and distances between

genotypes by multivariate statistics (i.e., principal component analysis and discriminant analysis), the methods involve analysis of agronomic traits which generally exhibit quantitative inheritance. The similarity or distance estimation is affected by the environment as well as by the number of traits included in the analysis. The data collection, if not available from yield trials, has a high cost and is time consuming. Gizlice *et al.* (1993a) evaluated 14 ancestral lines for GS using 10 different metric traits by principal component analysis. They pointed out high genetic similarity between CNS and S.100. They recommended that genetic similarities between parents in crossing should not be higher than 0.25, but they calculated 91 GS among the lines in their study and of the coefficients, only 21 were less than 0.25. Protein (isozyme) and molecular (RFLP) marker data were compared with multivariate analysis, and the authors concluded that multivariate analysis of quantitative traits could be more sensitive than markers in detecting genotypic differences when the number of molecular markers was low. These results similar to those obtained by Bhatt (1973). He reported that using multivariate techniques on parents as a predictor for variance in segregating populations in wheat showed promising results.

Cox *et al.* (1985) reported that Mandarin (Illinois) and Mandarin (Ottawa) were identical at 13 isozyme loci and 6 morphological marker loci. They stated that agreement between similarity estimates based on biochemical markers and quantitative traits was generally greater in self pollinated species than cross pollinated species.

As the third and latest method, DNA or molecular markers offers easy and quick measure of GS in plants. DNA markers may serve not only to estimate GS, but also to genotype fingerprint for paternity rights, support marker assisted selection, and facilitate genetic mapping and gene isolation. Allen (1993) pointed out that molecular markers could be used for different purposes in plant breeding including establishment of unique genetic identity and evaluation of genetic relationships among individual genotypes. In the last two decades, several DNA marker techniques were developed. Of the molecular markers, RFLP (Botstein *et al.* 1980), simple sequence repeats (SSR) or microsatellite (Tautz, 1988), random

amplified polymorphic DNA (RAPD; Williams *et al.* 1990), DNA amplification fingerprinting (DAF; Caetano-Anollés *et al.* 1991), and amplified fragment length polymorphism (AFLP; Vos *et al.* 1995) have commonly been employed in plant research. RFLP and SSR markers generally show codominant inheritance while RAPD, DAF and AFLP exhibit mostly dominant inheritance.

Akkaya *et al.* (1992) were the first discover SSR markers in soybean, found mainly $(AT)_n$ and $(ATT)_n$ repeats. CA and GT repeats were not a good source of polymorphism in soybeans in contrast to the human genome.

Shoemaker *et al.* (1992) compared cultivars Corsoy and Williams by RFLP markers. The genetic distance between Corsoy and Williams was 46% while it was 17 % between Corsoy and its parent Harosoy. They suggested that an extensive molecular pedigree analysis could determine which chromosomal segments were retained or lost during 50 years of selection and cultivar improvement.

Skorupska *et al.* (1993) evaluated southern 108 soybean genotypes, including some ancestral lines, to detect molecular diversity. They suggested that molecular polymorphism might be narrow within maturity groups (MG); However, they concluded that MG VI and VII cultivars have retained a high level of genetic variability.

Shoemaker *et al.* (1993) reported that frequency of an RFLP marker, K-418a, increased from 0.11 in first generation soybean cultivars to 0.30 in the fifth generation cultivars. They also stated that allele combinations which breeders consistently selected for or against could be determined by evaluating cultivars generation by generation.

Rongwen *et al.* (1995) assayed 96 soybean cultivars to investigate SSR marker polymorphisms for distinguishing genotypes. Ninety-one soybean lines were assayed by seven SSR loci. One of the primers, Sat43, showed 10 alleles in a single locus. They reported that SSR markers showed more polymorhism than RFLP markers; indeed 10 to 15 SSR markers sufficed to differentiate closely related soybean genotypes.

Lorenzen *et al.* (1995) attempted to trace soybean cultivar pedigrees by using RFLP

markers, and concluded data that the reported pedigrees of Lincoln and Ogden were most likely incorrect. They suggested that the ability to follow regions of chromosomes from parents to offspring through cultivar development generation should give insight into understanding what happened at the molecular level during soybean breeding over time.

Doldi *et al.* (1997) assayed 18 soybean genotypes with 33 RAPD and 12 SSR primers to evaluate genetic relationships. SSR data were in good agreement with pedigrees, but not RAPD data. They concluded that marker data were very useful and reliable to determine genetic relationships among cultivars in the absence of pedigree information.

Prabhu *et al.* (1997) estimated DAF-GS, RFLP-GS and CP using 10 soybean cultivars. These cultivars were analyzed by 53 RFLP probes and 12 DAF primers. Three GS estimates were significantly correlated. Clusters formed by these GS estimates were almost identical. They found that two selected polymorphic DAF primers were able to distinguish nine cultivars while five RFLP probes could separate 10 genotypes.

Sneller *et al.* (1997) conducted experiments in five locations with 31 plant introductions (PI), 11 populations (F_5 and F_6) derived from five northern cultivars x southern lines, 15 southern lines and 9 southern x southern populations. PIs, parents of crosses and 57 southern elite lines were screened by 27 RLFP probes. They were able to distinguish southern lines from northern lines by RLFP. GS by RFLP (RFLP-GS) within southern and northern material was similar. The correlations within groups between RFLP-GS, and yield, plant height, lodging and seed size were not significant.

Kisha *et al.* (1997) developed three sets of soybean populations: (1) eight $F_{2:3}$ populations, (2) 21 $F_{3:4}$ populations, and (3) 24 $F_{4:5}$ populations. The parents in the crosses were screened by 25 RFLP probes. RFLP-GS and CP were estimated. In the first set, CP and RFLP-GS were associated only with the variance of plant height. There was no relationship in the second set, and CP was correlated with the variance of plant height and maturity in the third set while RFLP-GS was correlated only with the variance of maturity. Yield was not associated with CP or with RFLP-GS.

Manjarrez-Sandoval *et al.* (1997) created 24 F_2 populations of soybeans to estimate heterosis and to predict heterosis by CP and RFLP-GS. The CP and RFLP-GS were very significantly correlated , but neither of them predicted heterosis for yield due to high genotype x environment interaction. However, heterosis for plant height was predicted by these GS estimates. They concluded that rather than GS estimates, extensive field testing is needed to identify favorable combinations in soybeans.

Kisha *et al.* (1998) classified 165 North American soybean lines into five groups. The groups were ancestral lines, northern elite lines, southern elite lines, PIs from MG I and II, and PIs from MG V to VII. Soybean lines were assayed by 53 RFLP markers. Clusters by RFLP data were consistent with their grouping with the exception of ancestral lines. RFLP-GS was associated with CP ($r= 0.61$). Genetic diversity in the northern soybean gene pool was greater than that in the southern soybean gene pool. Twenty ancestral lines contributed 77 % to the parentage of the elite gene pools.

Studies on GS estimates in other crops have been conducted. In wheat, Barret *et al.* (1998) evaluated 43 spring and winter wheat by 16 RFLP primers. Overall CP and RFLP-GS correlation was significant ($r= 0.42$). However, when cultivars were divided as winter and spring, then r was 0.45 and -0.02 for winter and spring wheat, respectively. However, Burkhamer *et al* (1998) got different results. They created 12 $F_{3:5}$ wheat populations by 10 parents. CP, AFLP-GS, and STS-GS (sequence-tagged sites) were interrelated as correlation changes from 0.65 to 0.86, but none of GS estimates was a good predictor of genetic variance in the progeny. In a more comprehensive study, Bohn *et al* (1998) estimated genetic diversity among 55 wheat cultivars by 117 RFLP probes, 16 AFLP primers and 21 SSR primer pairs. They also developed 30 $F_{4:6}$ or $F_{4:7}$, and tested 22 lines from each population in the field to get variances for quantitative traits including yield. Clusters formed by CP, AFLP-GS, RFLP-GS and SSR-GS did not reveal any common pattern. There was not any significant relationship with any GS estimate and variance for quantitative traits.

Russell *et al.* (1997) compared RFLP, AFLP, SSR and RAPD by determining genetic

relationships among 18 cultivated barley accessions. SSR showed the lowest GS estimates among markers techniques. Information content of SSRs was higher than the others due to high number of alleles per locus. SSR marker data were not correlated with other marker techniques. They concluded that SSR did not seem to be a suitable marker technique to assess genetic relationships among barley cultivars.

In corn (*Zea mays*), Smith *et al.* (1997) compared 131 SSR and 80 RFLP markers, and pedigrees. SSR and RFLP showed high correlations with CP (r =0.80 and 0.81respectively). SSR and RFLP data were very similar to each other, but they recommended to use SSR markers for genetic discrimination of maize lines due to higher reproducibility and polymorphism compared to RFLP. Similar results were obtained by Senior *et al.* (1998) in corn. Ninety-four elite maize inbred lines were evaluated by SSR markers and CP. The patterns established for genetic distances by SSR were consistent with pedigrees. They also concluded that SSRs were more efficient to measure genetic diversity than RFLP. Pejic *et al.* (1998) also found significant relationships among CP, RFLP-GS, RAPD-GS and SSR-GS. RAPD-GS yielded the lowest correlation, and the cluster formed by RAPD was not similar to the other three clusters by CP, RFLP and SSR. Clusters by molecular markers showed some discrepancies with pedigrees. SSR provided the highest level of discrimination among corn lines. Ajmone-Marsan *et al.* (1998) investigated genetic diversity by RFLP and AFLP and their association with hybrid performance in corn. Clusters by AFLP and RFLP were consistent with pedigree information. Genetic diversity estimates were correlated positively with F_1 performance for yield, but they were too small to be predictors.

Cultivar improvement is a long, tedious and costly process. If the variance and transgressive segregants for the trait(s) to be selected are predictable by GS or any other means, the number of crosses and field testing could be reduced considerably. In soybeans and other crops, research on GS and the methods to estimate it has been accelerated in the last two decades. This study was conducted to: (1) show genetic variability in soybean over time by DAF and SSR markers, and by agronomic traits, (2) evaluate associations of quantitative

trait variances in southern x southern and northern x southern crosses with DAF-GS, SSR-GS, quantitative trait-GS (QT-GS) and CP, and (3) compare DAF, SSR, QT-GS and CP for cultivar classification.

CHAPTER 2

Materials and Methods

Cultivars/Lines and Crosses

Southern and northern ancestral lines and cultivars, CNS, S.100, Ogden, Lee, Unknown (Hill), Essex, TN 5-85, Hutcheson, Mandarin (Ottawa), Harosoy, Williams, and BSR 101, were chosen for inclusion in this study (Table 4). The seeds were obtained from Dr. Randal L. Nelson, soybean curator, USDA, ARS, Urbana, Illinois. The twelve entries

Table 4. Selected southern and northern cultivars/lines, and their maturity group (MG), cultivar generation, release year and pedigree.

Cultivar/Line	Abbreviation	MG	Cultivar Generation	Release Year	Pedigree‡
Southern					
CNS	CN	VII	Ancestor	-	Introduction
S.100	S1	V	Ancestor	-	Introduction
Ogden	OG	VI	1	1953	Tokyo x PI 54610
Lee	LE	VI	1	1954	S.100 x CNS
Hill (Unknown)	UN	-	-	-	-
Essex	ES	V	3	1972	Lee x S5-7075
TN 5-85	TN	V	5	1986	Essex x D68-127
Hutcheson	HU	V	5	1987	Essex x V68-1034
Northern					
Mandarin+	MA	0	Ancestor	-	Introduction
Harosoy	HA	II	1	1951	Mandarin+ x AK Harrow
Williams	WI	III	3	1971	Wayne x L57-0034
BSR 101	BS	I	5	1988	A76-304020 x L69U40-16-4

+Mandarin (Ottawa)

‡ The complete pedigrees tracing back to ancestral lines are presented in Appendix 1 and 2.

represent ancestral lines, and selected first, third and fifth generation southern and northern cultivars representing different eras of production from the 1950's to the 1990's. The lines/cultivars have bee chosen based on three criteria; The ancestral lines CNS, S.100 and Mandarin were chosen because they are major contributing lines to the modern gene pool of ancestral lines (Delanay *at al.* 1983 and Gizlice *at al.* 1993a). The cultivars were chosen because they represent different generations of related cultivars and they were widely grown during their era of production. In addition, most of them have been utilized extensively as parents in both public and private breeding programs.

Crosses were made between the two southern ancestral lines between one southern and one northern ancestral lines (Table 5). Crosses were also made between first, third and fifth generation southern x southern or northern x southern cultivars (Table 5). Twenty F_1 seeds were sent to Costa Rica to produce F_2 seeds in winter nurseries. In addition, 50 seed from each F_1 plant were space-planted at Knoxville in 1997 in 3 m long rows in an incomplete block design. The plant spacing within rows was approximately 6 cm with 76 cm between rows. Approximately 250 plants from each F_2 population and 50 plants from each parent were selected randomly at R1 stage and tagged. Plant height, maturity, protein and oil content, and yield were recorded on selected plants.

Seeds from selected F_2 plants were planted to grow $F_{2:3}$ lines at Knoxville in 1998 in 3m long rows in an incomplete block design. $F_{2:3}$ lines from each populations were equally represented in each block and 12 parents was randomized with the lines in each block. The parents and $F_{2:3}$ lines in each block were randomized randomly within the block by SAS procedure SORT (SAS Institute, Inc. 1998, Cary, NC).Parent lines were placed in each block to estimate the experimental error as accurately as possible. The spacing within the row and between the rows was approximately 4 cm and 76 cm, respectively. At harvest, 50 cm from the each end of rows was discarded as border effect. The traits recorded on F_2 plants were also recorded in each row.

Table 5.The list of crosses made between southern x southern and northern x southern soybean cultivars/lines

Cross	Cultivar generation	Type of cross
S.100 x CNS	Ancestral lines	Southern x southern
Mandarin x CNS	Ancestral lines	Northern x southern
Lee x Ogden	First	Southern x southern
Harosoy x Lee	First	Northern x southern
Essex x Unknown	Third	Southern x southern
Williams x Essex	Third	Northern x southern
TN 5-85 x Hutcheson	Fifth	Southern x southern
BSR 101 x TN 5-85	Fifth	Northern x southern

DNA Analysis

Genomic DNA was isolated from very young leaves of parents according to Dellaporta *et al* (1983). DNA concentration measured in a TK0100 Dedicated Mini-Flourometer (Hoeffer Scientific Instruments, San Francisco, CA) by following manufacturer instructions.

Fifteen SSR soybean primers (Table 6) were chosen by considering linkage groups to broadly cover the soybean genome as much as possible. PCR reaction mixes and PCR protocol were adjusted according to Rongwen *et al* (1995), but instead of 30 µl volume, 10 µl reaction mixes were employed for each sample. Primer pairs were synthesized by Research Genetics (Huntsville, AL). Detection and visualization of markers were fulfilled as in DAF markers, but gels were run for 3.5-4 h at 350 W using a Protean gel chamber (Bio-Rad Inc., Richmond, CA). In silver staining fixation and staining stages were increased up to 15 and 30 min, respectively. After gels were dried, the size of bands (markers) were estimated using an AlphaImager 2000 (Alpha Innotech Corp., San Leandro, CA). AlphaImager results were

Table 6. The list of SSR primers, primer sequence, and linkage group of SSR markers

Primer	Sequence		Linkage group (USDA/ARS map)
Sct001	Forward	TTAGTTTCCCTCTCTCTCT	J
	Reverse	TGTTCCTTCGCTCAC	
Sat003	Forward	TGATTTTTGGTGTAGAACTC	M
	Reverse	AAATTGGTTAGCTTACTCCA	
Sat069	Forward	GACCAGCTGAAGAAA	D1b
	Reverse	GAATACCCATCATTACTTAA	
Sat085	Forward	GTTTTAGATCCTTAAATTTGT	C1
	Reverse	GGAAGCAAGTAGCT	
Satt006	Forward	AACCCTTTAAATTCAAATT	L
	Reverse	AAAGCGCATAAACAATAT	
Satt030	Forward	AAAAGTGTTAACCAAGCC	F
	Reverse	TTAAATCTTATGTTGATGC	
Satt050	Forward	AACGTCCCAACATTG	A1
	Reverse	CTTGTAAACATATAGG	
Satt052	Forward	AATAAAATTAGGATAAGTGATAAG	H
	Reverse	AGAAAAAAGAAAATGTCA	
Satt079	Forward	GTCGAAGATACACAATTAGAT	C2
	Reverse	TTTAGACACAAATTTATCACT	
Satt089	Forward	AAAATTAACGGTTGAGAAT	A2
	Reverse	TAAGGAAATAAAAGAAGAAAAG	
Satt175	Forward	ACCTCGCTCTCTGTTTCTCAT	M
	Reverse	TGACCACCCCTATTCCTTAT	
Satt184	Forward	CGCTATGTAGATTTCAAATCGC	D1a
	Reverse	CACTTACTGTTACTAT	
Satt197	Forward	ACTGCTTTTTCCCCTCTCT	B1
	Reverse	GATACCCCCAACATTATTTGTAA	
SoyHSP176	Forward	TTTTGTTTAAGTTACTGTACTGT	F
	Reverse	TAGTCTTCTACAACCTTATA	
SoyRPRP1	Forward	GTGCCAAATTACATCA	J
	Reverse	ATGGGAACAAGTACATAA	

compared with sequencing gel estimation and the results were in good agreement.

A total of 15 DAF primers (Table 7) were chosen randomly from available primers at the University of Tennessee, Plant Molecular Genetics Laboratory, Knoxville, TN. Reaction mixes for PCR contained 2 ng of genomic DNA, 3.5 mM Mg^{2+}, 3 μM primer, 200 μM of each nucleotide, 1x PCR Buffer containing 50 mM KCl and 10 mM Tris-HCl, and 1 unit Stoffel Taq DNA Polymerase (Perkin Elmer, Norwalk, CT) in 10 μl total volume. PCR was run 36 cycles; denaturation at 95° C for 30 s, annealing at 55°C for 2 min, and extension at 72°C for 30 s in a MJ Reseach PTC-200 DNA Engine (MJ Research, Watertown, MA). PCR products were separated on a 5% polyacrylamide gel containing 7 M urea. Gels were run for approximately 60-65 min at 100 W constant power using a mini-Protean II gel chamber (Bio-Rad Inc., Richmond, CA). The markers were detected and visualized by silver staining as described by Bassam *at al.* (1991). The markers between 200 and 700 bp were considered, and weak bands were ignored. Each DNA sample was repeated twice.

Table 7. The list of DAF primers and primer sequence

Primer	Sequence (5'-3')	Primer	Sequence (5'-3')
A10	GCCCGCGT	A52	GCCCGACA
A13	GCCCGCAC	A57	GCCCGAGC
A16	GCCCGCAA	B10	GCAGGCGT
A19	GCCCGTCG	B11	GCAGGCGG
A34	GCCCGGCT	8-4	GTATCGCC
A36	GCCCGGCA	8-26	CCAGGTGG
A37	GCCCGGTC	8-45	GGACCCGC
A41	GCCCGGGC		

Statistical Analysis

GS estimates between cultivar/line pairs were accomplished by considering only polymorphic markers for DAF and SSR. Therefore, we modified Nei and Li's (1979) genetic distance which is $2n_{xy}/n_x+n_y$ where n_{xy} is the number of bands different in two cultivars, n_x is the number of bands in cultivar x and n_y is the number of bands in cultivar y. This genetic distance is useful when the number of bands in two genotypes are not the same. Therefore, when only polymorphic bands are considered, it can be modified n_{xy}/n_t where n_t is the total number of bands. GS is 1 minus genetic distance. Non-polymorphic markers were ignored, because it would cause DAF to lead to higher GS estimates compared to SSR. Covariances between genotypes were estimated by SAS PROC INBREED (SAS Institute, Cary, NC), then covariance was divided by two to get CP. Multivariate agronomic distances were obtained by SAS Proc DISCRIM using agronomic data. Pairwise generalized squared distances (D^2) in discriminant analysis were taken as Quantitative trait-genetic distance (QT-GD) which was:

$$QT\text{-}GD = (\bar{x}_x - \bar{x}_y)' \, \mathrm{cov}^{-1} \, (\bar{x}_x - \bar{x}_y)$$

where $\bar{x}_x$ and $\bar{x}_y$ were the overall means of cultivars x and y. In other words, QT-GD was the sum of squares of the differences between two genotypes weighted by the covariances among the differences. QT-GD were converted to QT-GS by setting largest distance (D^2_{max}) to no similarity, and using the largest distance as the denominator for QT-GD (D^2_g) as in the following formula (Gizlice *et al*, 1993a):

$$QT\text{-}GS = 1 - (D^2_g / D^2_{max})$$

Broad sense heritabilities (h^2) or degree of genetic determination which is $\sigma^2_{genotype}/\sigma^2_{phenotype}$ for F_2 plants were computed by using variance components obtained by SAS PROC MIXED (SAS Institute, Inc. 1998, Cary, NC). The statistical model was $y=\mu$ *block genotypes error* where y is the agronomic trait, and μ is the general mean. Genotypes (F_2 plants and parents) was treated as random variables. A reduced model was used in mixed model to estimate variance components. In other words, when block effect was not significant, it was dropped from the model.

Correlations among GS estimates were obtained by SAS PROC CORR. Clusters were formed by average linkage methods in SAS PROC CLUSTER. There is no parameter to compare different cluster methods; therefore, the methods can be compared based on other available data such as pedigrees and phenological observations. A method which gives more meaningful results compared to the others should be preferred. We compared average linkage, complete linkage, centroid, and Ward's minimum variance methods. Clusters by average linkage method seemed to be more appropriate than others. When different GS estimates were used together for clustering, the data were standardized, because any trait which has larger values dominates the cluster (Milligan and Cooper, 1987).

CHAPTER 3

Results and Discussion

Genetic Trends in Agronomic Traits, and Molecular Marker Diversity

Heritability

Broad sense heritabilities over cultivar generations may contribute to our understanding of what kind of genetic trends occurred in the southern U.S., and what was the potential of northern x southern crosses for any given cultivar generation. Heritability was the highest for maturity and the lowest for oil content. Plant height and yield had high heritabilities, but protein content had a low heritability (Table 8 and Appendix 3).

Broad sense heritabilities for yield showed the decline in genetic variation for yield. In southern x southern crosses, heritability for yield decreased 14 % from S1 x CN (ancestors) to TN x HU (fifth generation; Figure 5). However, this decrease in northern x southern crosses was more apparent (38 %). Higher heritabilities were estimated for southern x southern crosses than for northern x southern crosses. Heritabilities of the fifth generation crosses between TN and HU, and BS and TN were 0.59 and 0.50 (Table 8). In addition to broader yield variation in southern x southern crosses compared to northern x southern crosses, the mean of F_2 populations in the first and fifth generation crosses was higher in southern x southern crosses than northern x southern crosses. However, it was similar in ancestral line and third generation crosses (Figure 6 and 7). The genetic variance in southern soybean declined only 9.3 % from ancestral lines to the fifth generation (Table 8). Decline in genetic variation for yield was expected because of repeated selection for yield over time in soybeans. Similar results were also obtained or assumed by St. Martin (1982), Burton (1987), Gizlice *et al.* (1994) and Kisha *et al.* (1998). Higher genetic variation in southern x southern populations can be explained by better fitness of southern material to Knoxville

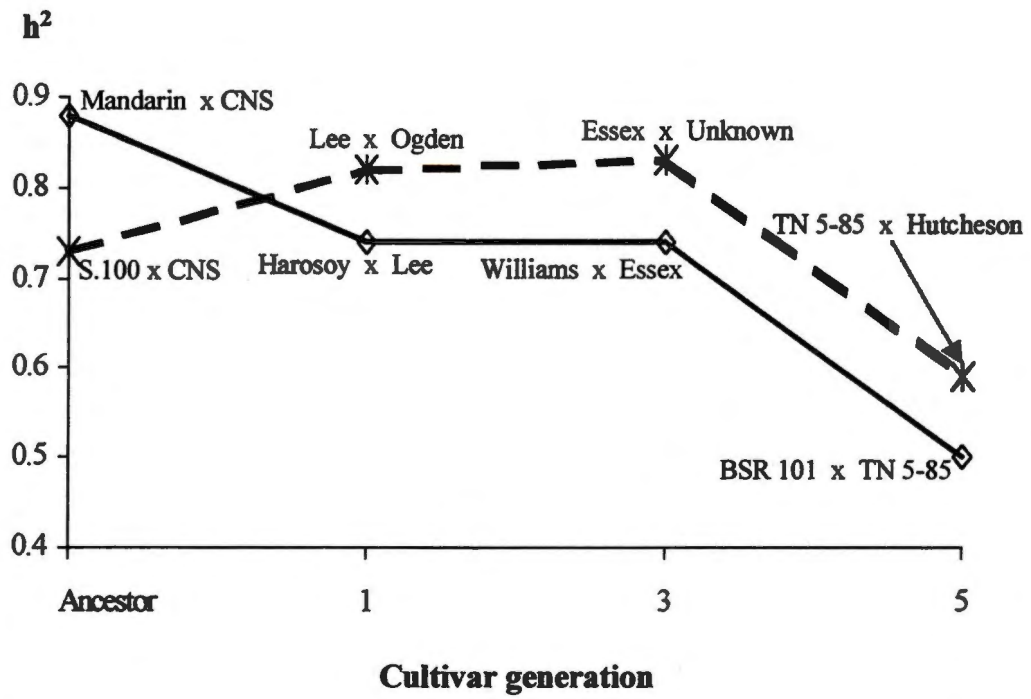


Figure 5. Heritability changes for yield over cultivar generation in specific F_2 populations of southern x southern and northern x southern soybean crosses

conditions, so their variation might be exploited better than northern x southern crosses. This result indicates that there is still remaining variation in southern soybean gene pool to gain yield increase by selection. On the other hand, decline in genetic diversity is a warning to increase it by either making crosses with northern lines or plant introductions.

Soybean protein and oil content have been targets for selection following yield. The negative relationship between protein and oil content not only makes it difficult to increase both of them in the same genotype, but also selection for one causes reduction in variability for both of them. Variability for oil content was decreased 40 % over time in southern populations while it did not change in northern x southern populations (Figure 8). It seems that variability for oil content decreased to a level that genetic gain may be limited by crossing recent southern cultivars (Table 8). Southern breeders need to make crosses with northern lines and PIs from MG IV through VII to increase variability for protein and oil content.

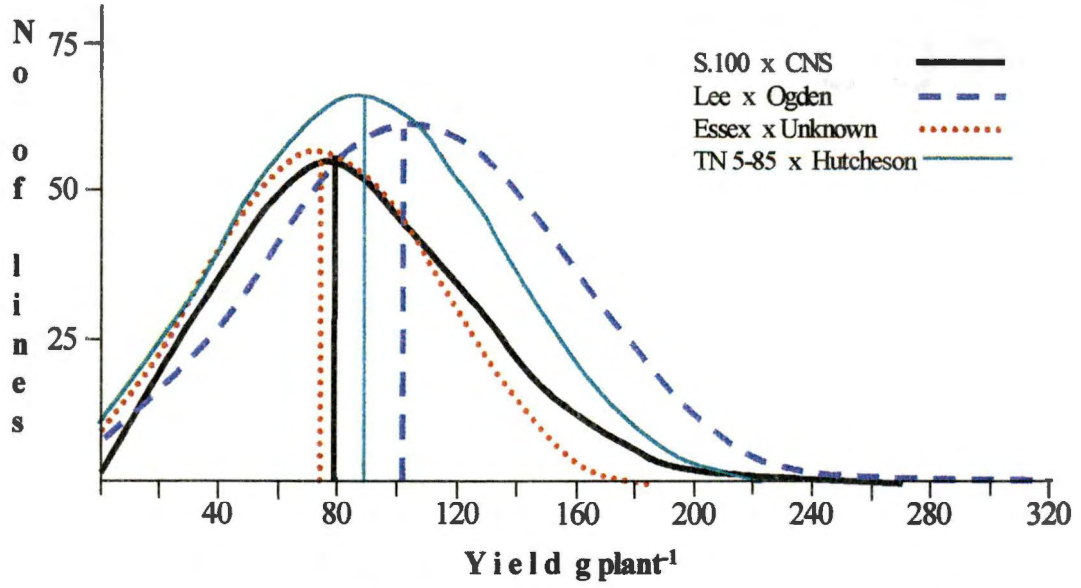


Figure 6. The distribution of yield and population mean in southern x southern soybean F_2 populations

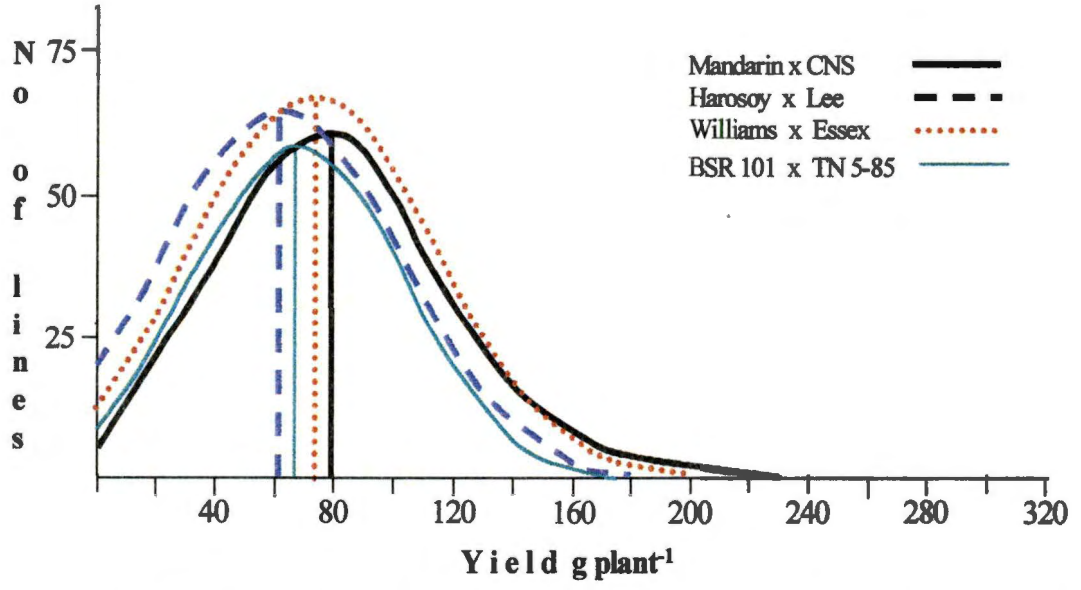


Figure 7. The distribution of yield and population mean in northern x southern soybean F_2 populations

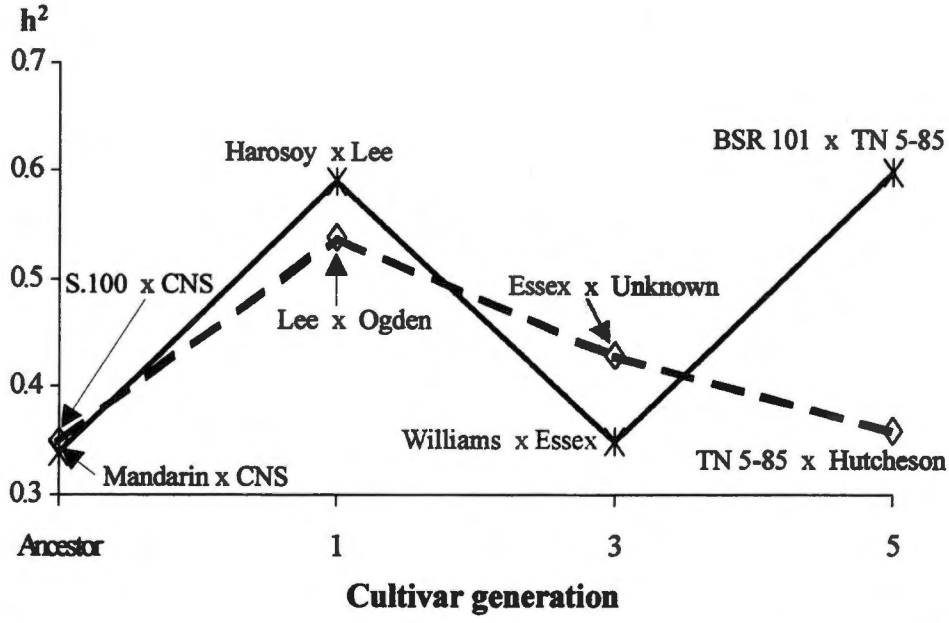


Figure 8. Heritability changes for oil content over cultivar generation in specific F_2 populations of north x south and south x south soybean crosses

The northern x southern crosses showed higher h^2 than southern x southern crosses for maturity. Heritability for maturity across the cultivar generations was similar (Appendix 3). The heritability range in all eight populations changed from 0.67 (S1 x CN) to 0.97 (BS x TN and MA x CN). Heritability for plant height showed a decrease over time in southern x southern and northern x southern crosses. The lowest (0.45) and highest h^2 (0.89) were estimated in TN x HU and MA x CN F_2 populations, respectively. It shows that the remaining variation even within southern gene pool seems quite sufficient to make progress for this trait in the future.

Table 8. Genotypic variance (σ^2_g), phenotypic variance (σ^2_p) and broad sense heritability (h^2) for seed yield, protein content and oil content in different F_2 populations of soybean

Population	σ^2_g	σ^2_p	h^2	Population	σ^2_g	σ^2_p	h^2
Yield							
Southern x southern				**Northern x southern**			
S.100 x CNS	1467.7	2006.3	0.73	Mandarin x CNS	2915.4	3315.6	0.50
Lee x Ogden	2467.8	3004.0	0.82	Harosoy x Lee	1005.9	1363.0	0.65
Essex x Unknown	1331.6	1599.7	0.83	Williams x Essex	1309.3	1765.6	0.13
TN 5-85 x Hutcheson	1331.9	2256.0	0.59	BSR 101 x TN 5-85	682.2	1370.6	0.67
Protein content							
Southern x southern				**Northern x southern**			
S.100 x CNS	0.933	2.687	0.35	Mandarin x CNS	0.911	1.834	0.50
Lee x Ogden	1.051	1.945	0.53	Harosoy x Lee	1.601	2.451	0.65
Essex x Unknown	0.770	1.203	0.64	Williams x Essex	0.169	1.285	0.13
TN 5-85 x Hutcheson	0.302	1.186	0.25	BSR 101 x TN 5-85	1.572	2.334	0.67
Oil content							
Southern x southern				**Northern x southern**			
S.100 x CNS	0.153	0.435	0.35	Mandarin x CNS	0.133	0.389	0.34
Lee x Ogden	0.158	0.292	0.54	Harosoy x Lee	0.264	0.445	0.59
Essex x Unknown	0.089	0.208	0.43	Williams x Essex	0.084	0.241	0.35
TN 5-85 x Hutcheson	0.092	0.258	0.36	BSR 101 x TN 5-85	0.240	0.397	0.60

Molecular markers

SSR marker diversity (Figure 9) within southern material was identical to that between northern and southern cultivar/lines until the third generation cultivars. Identical results until the third generation could be caused by low number of markers and coincidence of cultivars showing this result. However, the point in Figure 11 is the decrease of cultivars showing this result. However, the point in Figure 11 is the decrease of SSR diversity between

showing this result. However, the point in Figure 11 is the decrease of SSR diversity between both southern and southern, and northern and southern genotypes following the third cultivar generation. In addition, the decrease between southern and southern accessions was more tragic. From the third generation to the fifth generation, the decrease was 55 and 18 % between southern and southern and northern and southern genotypes, respectively. The number of polymorphic markers between BS and TN, and TN and HU were 5 and 9 out of 15 markers, respectively.

The cultivar "Hill" was crossed with Essex and it was observed by examining the parent rows of Hill in F_2 generation and other Hill plots in yield trials that the genotype was not Hill. The Hill from University of TN soybean breeding program and from Stoneville, MS (breeder institution) were compared with the genotype. Seven out of 15 SSR markers showed that the genotype was not Hill (Figure 9). Therefore, it is called "Unknown"in this research.

DAF marker diversity (Figure 10) between cultivars/lines declined between southern cultivars following the first generation cultivars (Figure12). The diversity between northern and southern genotypes did not show a declining trend. Furthermore, it was increased between fifth generation cultivars, TN 5-85 and BSR 101. Mandarin and CNS showed differences for 28 markers out of 29, and two southern ancestral lines exhibited differences for 21 markers. The highest number of polymorphic DAF markers in southern material was found between Lee and Ogden.

St. Martin (1982), Delaney *et al.* (1983), Allen and Bhardwaj (1987), Gizlice *et al.* (1993a and 1994) pointed out the use of a limited number of ancestors repeatedly in soybean breeding programs. Broad sense heritabilities and molecular marker diversity by SSR and DAF evidenced that the repeated use of same germplasm started to be an issue related to the genetic diversity after the 1970s (third generation cultivars). Marker diversity by SSR and DAF suggests that this problem may be overcome by integration of northern and southern lines.

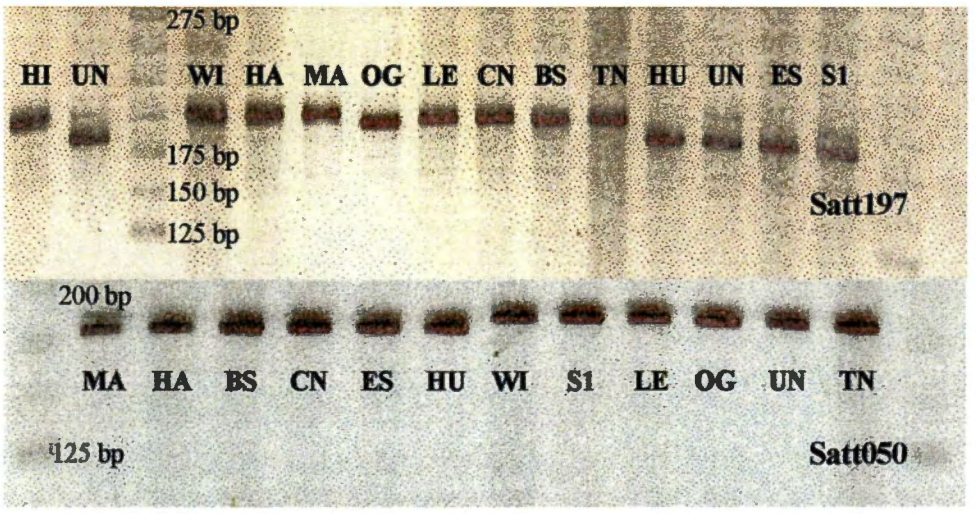


Figure 9. The Satt197 and Satt050 SSR markers showing differences among12 soybean genotypes.HI = Hill, MA= Mandarin, HA= Harosoy, WI= Williams, BS= BSR 101, CN= CNS, S1= S.100, LE= Lee, OG= Ogden, ES= Essex, UN= Unknown, HU= Hutcheson, TN= TN 5-85

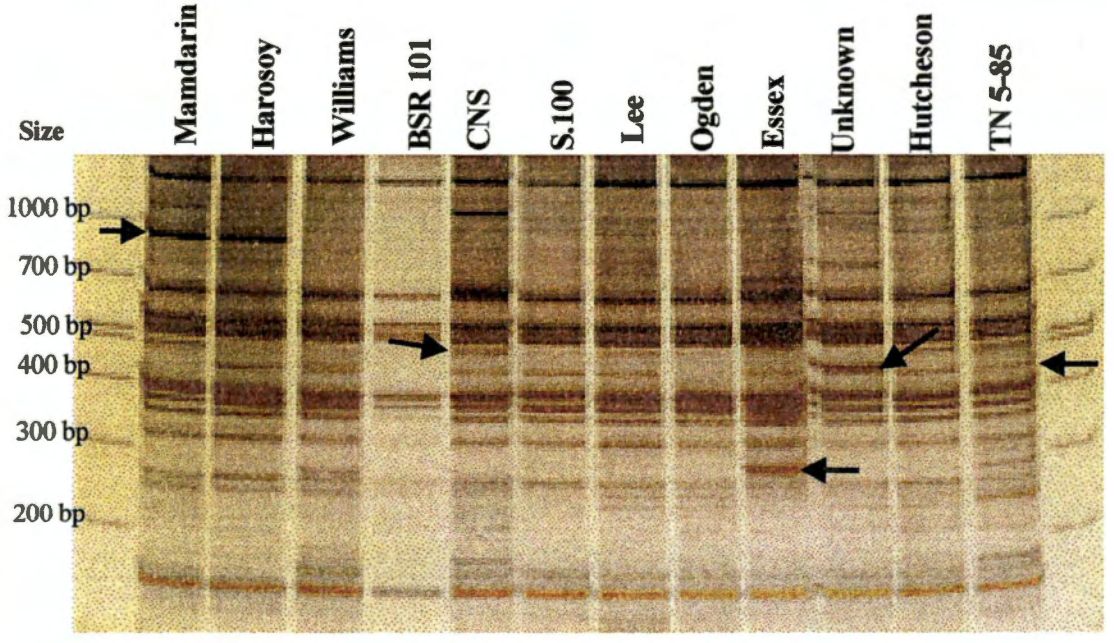


Figure 10. DAF primer A10 showing polymorphisms among soybean cultivars/lines from different cultivar generations.

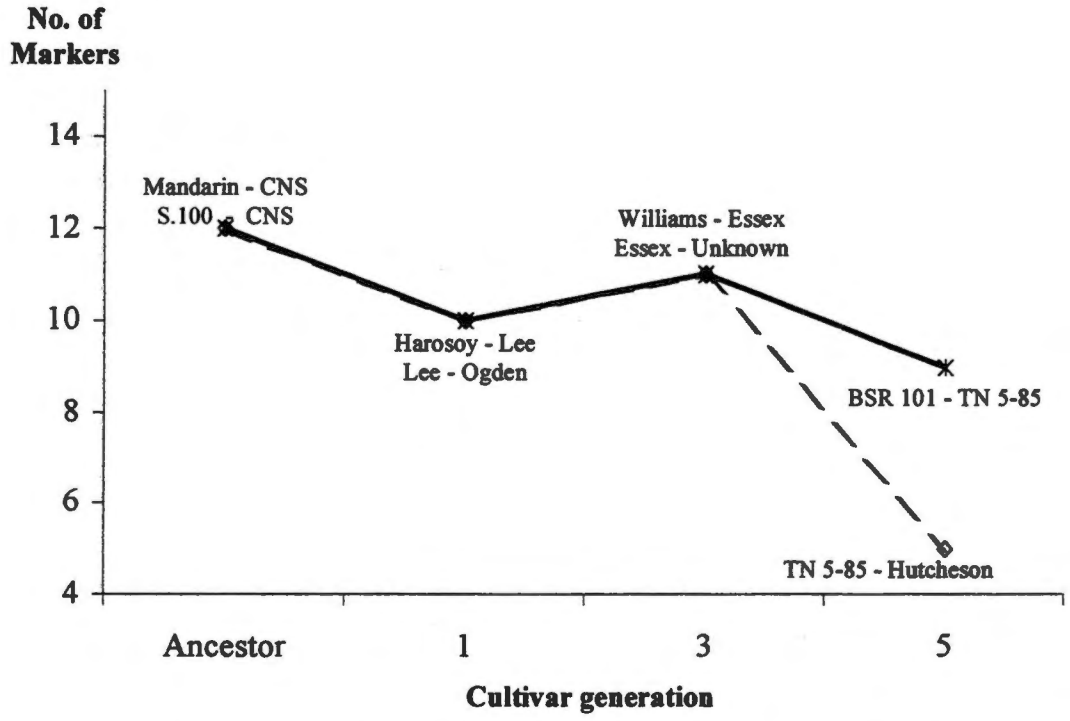


Figure 11. Number of SSR markers which show polymorphism between soybean cultivars or lines from different cultivar generations

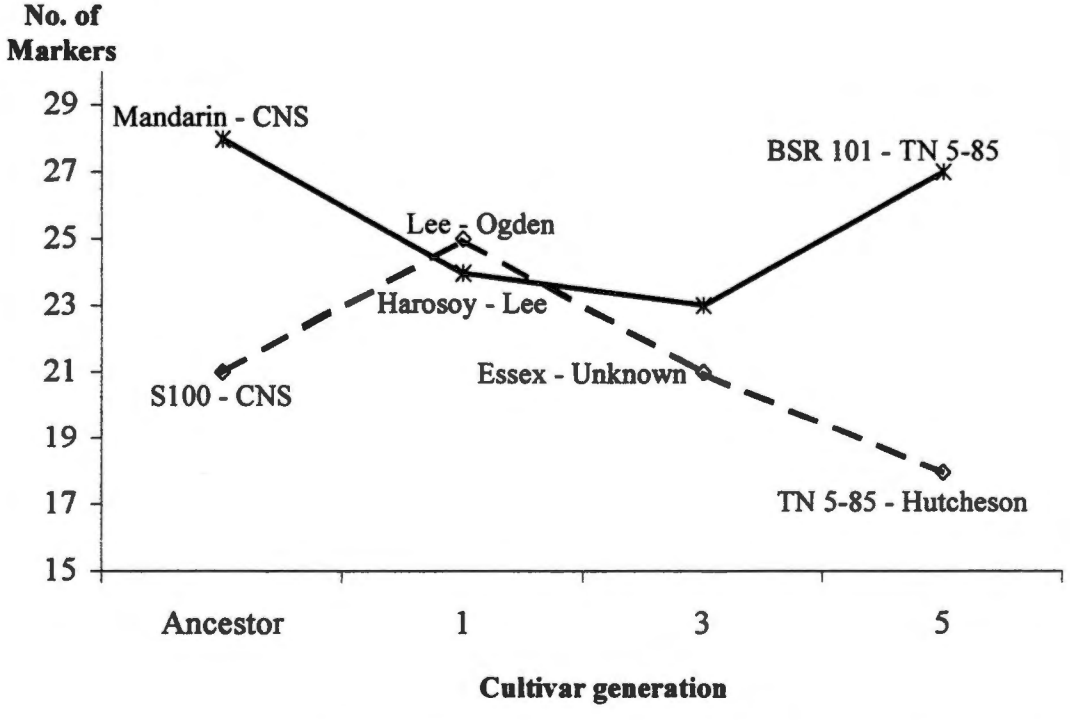


Figure 12. Number of DAF markers which show polymorphism between soybean cultivars or lines from different cultivar generations

Genetic Similarities and Their Associations with Agronomic Traits

When the parents are distant from each other, they should theoretically exhibit a large variation in the cross. Also, the probability of having transgressive segregants should be higher in the crosses between distantly related parents than in the crosses made between closely related parents. However, this theoretical assumption is difficult to demonstrate in self pollinated crops. Better predictors than agronomic traits and CP are needed to choose suitable parents. It has been assumed that molecular makers will improve the predictable outcome.

The highest and the lowest GS estimates among cultivars were obtained by QT-GS and CP, respectively. DAF-GS were usually higher than SSR-GS (Table 9, 10, and Appendix 4, 5). TN 5-85 and Hutcheson, and Ogden and Lee were very similar to each other according to QT-GS (0.998), while CNS and Mandarin were the most distant lines. The CP indicated that TN 5-85 and Essex were the most closely related cultivars. This was supported by SSR-GS and DAF-GS. The highest similarity (0.67) by SSR-GS was found between S.100 and Lee, Essex and Hutcheson, and Essex and TN 5-85. On the other hand, it was the lowest between Harosoy and CNS. DAF-GS pointed out that S.100 and Lee were the closest genotypes, and BSR101 and Lee were the most distant cultivars. The CP was zero for 34 pairs of cultivars/lines out of 68 due to assumption that it is zero between ancestral lines for which pedigrees are unknown. Therefore, CP was highly biased in our research. Gizlice *et al.* (1993a) stated that this bias in soybean CP estimation was minimized for the cultivars released after 1980s. They also found that S.100 and CNS had high similarity based on agronomic traits. However, none of the GS estimates in this study confirmed such similarity. The source of conflict for QT-GS may come from different experimental conditions. They measured traits in phytotron controlled-environment facility which minimizes the effect of maturity in other traits.

It seems that SSR compared to DAF, CP and agronomic traits measured GS more efficiently for discrimination among genotypes. Nevertheless, SSR primers have been developed for polymorphism, and they become available after polymorphism is proven. Consequently, SSR markers underestimate genetic similarities. Genetic similarity estimation

Table 9. Genetic similarities estimated by SSR markers among soybean genotypes

	Cultivars											
	MA	HA	WI	BS	CN	S1	LE	OG	ES	UN	HU	TN
MA	1.00											
HA	0.53	1.00										
WI	0.33	0.27	1.00									
BS	0.20	0.53	0.33	1.00								
CN	0.07	0.00	0.33	0.27	1.00							
S1	0.27	0.20	0.47	0.27	0.20	1.00						
LE	0.07	0.27	0.27	0.33	0.53	0.67	1.00					
OG	0.20	0.20	0.27	0.20	0.13	0.27	0.27	1.00				
ES	0.13	0.27	0.33	0.20	0.20	0.27	0.27	0.33	1.00			
UN	0.20	0.20	0.27	0.33	0.20	0.40	0.20	0.33	0.27	1.00		
HU	0.13	0.27	0.40	0.33	0.27	0.40	0.33	0.27	0.67	0.20	1.00	
TN	0.07	0.27	0.33	0.40	0.27	0.47	0.53	0.53	0.67	0.20	0.53	1.00

MA= Mandarin, HA= Harosoy, WI= Williams, BS= BSR 101, CN= CNS, S1= S.100, LE= Lee, OG= Ogden, ES= Essex, UN= Unknown, HU= Hutcheson, TN= TN 5-85

by DAF seems undoubtedly better than agronomic traits and CP. Better comparison between CP and molecular markers can be done in an experiment which comprises the cultivars released in the last 10-15 years from southern and northern states. Gizlice *et al.* (1993a) suggested that QT-GS might be more efficient than molecular markers to detect differences among genotypes in the case of low number of markers. When the genotypes are tested in a very controlled conditions such as phytotron controlled environment facilities, the reliability of OT-GS will increase, and OT-GS may be efficient to reveal differences among cultivars/lines. In field conditions, the effect of environment and experimental error can be so high that they cover the differences. As a result, QT-GS could be less efficient to find differences among breeding lines compared to molecular markers as it happened in this study.

Table 10. Genetic similarities estimated by DAF markers among soybean genotypes

	Cultivars											
	MA	HA	WI	BS	CN	S1	LE	OG	ES	UN	HU	TN
MA	1.00											
HA	0.88	1.00										
WI	0.59	0.61	1.00									
BS	0.58	0.53	0.61	1.00								
CN	0.53	0.54	0.64	0.47	1.00							
S1	0.51	0.59	0.61	0.39	0.64	1.00						
LE	0.54	0.59	0.61	0.36	0.64	0.90	1.00					
OG	0.59	0.54	0.47	0.47	0.53	0.54	0.58	1.00				
ES	0.54	0.53	0.61	0.56	0.68	0.69	0.69	0.58	1.00			
UN	0.39	0.47	0.54	0.54	0.56	0.58	0.58	0.56	0.64	1.00		
HU	0.49	0.54	0.63	0.61	0.59	0.61	0.58	0.64	0.64	0.69	1.00	
TN	0.53	0.51	0.53	0.54	0.63	0.64	0.68	0.49	0.69	0.59	0.69	1.00

MA= Mandarin, HA= Harosoy, WI= Williams, BS= BSR 101, CN= CNS, S1= S.100, LE= Lee, OG= Ogden, ES= Essex, UN= Unknown, HU= Hutcheson, TN= TN 5-85

Correlations between GS estimates and broad sense heritabilities of five traits were estimated (Table 11). Significant and negative correlations were obtained between SSR-GS and h^2 for yield, CP and h^2 for plant height. DAF-GS were on the border for significance with h^2 for maturity. These association should be taken with caution, because sample size was very low (n=8), and data for traits were obtained from single F_2 plants. The significant and negative correlation between SSR-GS and yield heritability can be very useful in planning of crosses in soybean breeding programs. This result needs to be confirmed with more populations in several years and locations. The correlation implies that there will be more genotypic variation for yield when parents which show less SSR similarity. Previous studies did not give consistent results related to marker-variance associations (Sneller *et al.* 1997, Kisha *et al.*

Table 11. Correlation coefficients between genetic similarity estimates and broad sense heritabilities in eight F_2 populations in soybeans

Trait	DAF-GS	SSR-GS	QT-GS	CP
Maturity	-0.621^{+}	-0.103	-0.436	-0.082
Plant height	-0.573	-0.595	-0.572	-0.722*
Protein content	-0.503	-0.252	-0.422	-0.457
Oil content	-0.426	-0.185	-0.002	-0.300
Yield	-0.147	-0.790*	-0.305	-0.478

+, * correlation different from zero ($P \leq 0.10$ and $P \leq 0.05$, respectively)

1997, and Manjarrez-Sandoval *et al.* 1997).

QT-GS were assumed as a predictor for the variance in wheat (Bhatt, 1973). In crops, quantitative traits are regulated by genetic and environmental factors; thus they are affected by years, location, plot size, number of replication, sampling and so on. Therefore, estimation changes from experiment to experiment. Still, breeders take notes in their experiments and get an idea of diversity or similarity without GS estimation, and use the information in crossing programs. In some plant breeding programs, the CP is used together with field observations in making decision for parents by breeder; however Helms *et al.* (1995) found that CP was not a good predictor of cross variance in soybeans.

Almost all studies have chosen markers randomly to determine agronomic trait-marker associations. Rather than randomly selected markers, it seems that markers which show QTLs (Quantitative Trait Loci) for agronomic traits may serve better for this purpose. The GS obtained by this type of markers may be more helpful than CP to choose parents effectively in self pollinated crops.

Classification of genotypes and comparison of similarity estimates

It is worthwhile to discriminate cultivars for two reasons; (1) to give "fingerprint" to cultivars for patent rights, and (2) to be able to classify genotypes. The latter might be useful and practical in planning of crosses. Agronomic traits are under the effects of environmental factors. Coefficient of parentage may not give desired results due to known assumptions such as no relationship between unrelated individuals. Molecular markers, which are usually not affected by environmental factors, might offer a more accurate discrimination and classification of cultivars.

The higher the GS by any estimator means the less efficient discrimination. SSR markers showed the least GS among cultivars (Table 9 and Figure 13). In other words, they indicated the greatest discrimination among cultivars/lines. DAF markers were much better than agronomic traits to distinguish cultivars, but less efficient than SSR markers (Table 10 and Figure 14). QT-GS almost failed to differentiate 23 % (16 out of 68) of the cultivar pairs (Appendix 4). In these 16 instances, the QT-GS was higher than 0.97. Coefficient of parentage should not be compared with others for this purpose in this study because of having 50 % zero coefficient (Appendix 5).

Cluster analysis is a commonly used method to visualize relationships among genotypes. Although there are better statistics like principal component analysis and discriminant analysis to classify genotypes, molecular marker data is not applicable to these multivariate statistics due to absence of replication in molecular data. Soybean in the U.S. is grown in two distinct zones, northern and southern cultivars bred for these zones have been developed primarily from a northern soybean germplasm pool and a southern germplasm pool. SSR and DAF markers, CP and agronomic traits confirmed the divergence of soybean genotypes in two clusters; one having northern genotypes and the other southern genotypes (Figure 13,14, 15 and 16). Luedders (1977) assumed that there was not a distinction between southern and northern soybean lines, because Ogden was used as a parent in both north and south. Our study and Sneller *et al.* (1997) demonstrate that northern and southern soybeans are distant genetically.

Williams and the Unknown genotype were the sources of conflicts among DAF, SSR and agronomic traits. DAF markers placed the Unknown within southern cultivars although SSR and agronomic data showed it in the cluster of northern cultivars. On the other hand, SSR and agronomic data classified Williams within southern cultivars, but DAF markers placed Williams in the cluster for northern accessions. Coefficient of parentage might have given the best representation if it has not included Ogden in the northern cultivars. The Unknown genotype was also shown within northern accessions. However, the Unknown had zero CP with all genotypes; therefore it was expected to form a separate cluster.

CNS was involved in the pedigree of Williams. Also, Williams was more similar to southern soybeans than other northern genotypes in the sample. Thus, it was possible to see it within southern cultivars. When the clusters were examined relative to pedigree information, SSR markers were not able to show the relationship clearly between Lee and its parents CNS or S.100 (Figure 13). However, SSR-GS captured the similarity between Lee and its parents (Table 9). This inconsistency between GS estimation and cluster analysis indicates that cluster analysis may cause misgrouping of genotypes. DAF markers suggested that Lee resembled S.100 more than CNS. On the other hand, it is difficult to understand the cluster formed by Hutcheson and Ogden by DAF markers. SSR, DAF and agronomic data were pooled and clustered (Figure 17). Combined data gave a consistent grouping with pedigrees and field observations. To classify the cultivars and lines more accurately by cluster analysis, it is better to combine molecular marker data with agronomic observations.

The relationship between different estimates can help to decide on GS estimator. SSR had higher correlation coefficients than DAF with QT-GS (Table 12). It is consistent with the result that yield heritability was associated with SSR-GS. The lowest correlations were between QT-GS, and DAF-GS and CP. SSR and DAF markers showed similar and strong association with the CP. Therefore, we assume that CP is still useful in plant breeding although the assumptions can not be met. There is not any previous study which included SSR and DAF in the same experiment. As a dominant marker, DAF should generate similar pattern to AFLP where number of bands per lane is very high. Significant correlations among CP and

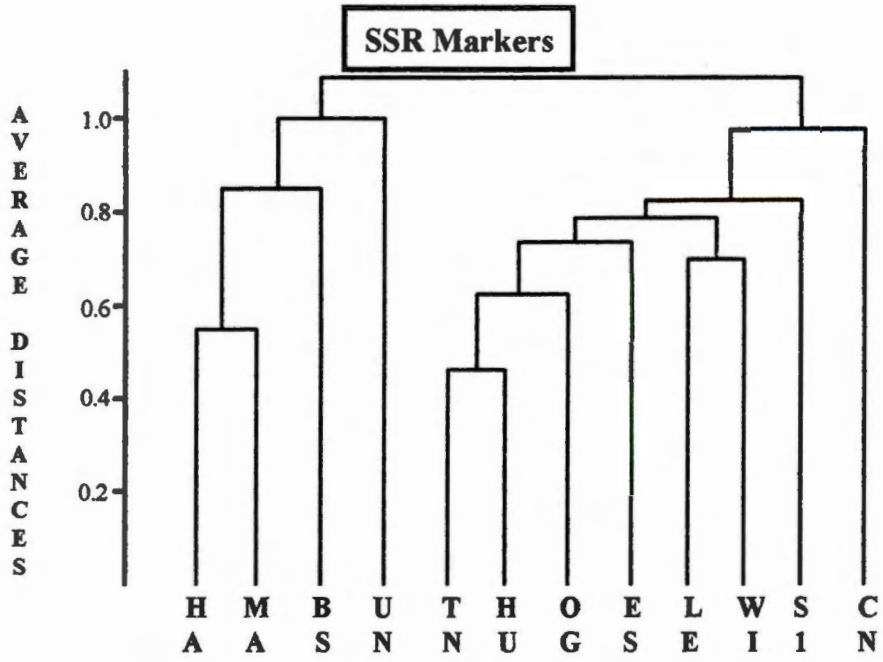


Figure 13. Clusters by SSR markers for northern and southern soybean cultivars/lines representing different eras. BS= BSR 101, WI= Williams, MA= Mandarin, HA= Harosoy, UN= Unknown, OG= Ogden, HU= Hutcheson, CN= CNS, TN= TN 5-85 ES= Essex, S1= S.100, LE= Lee

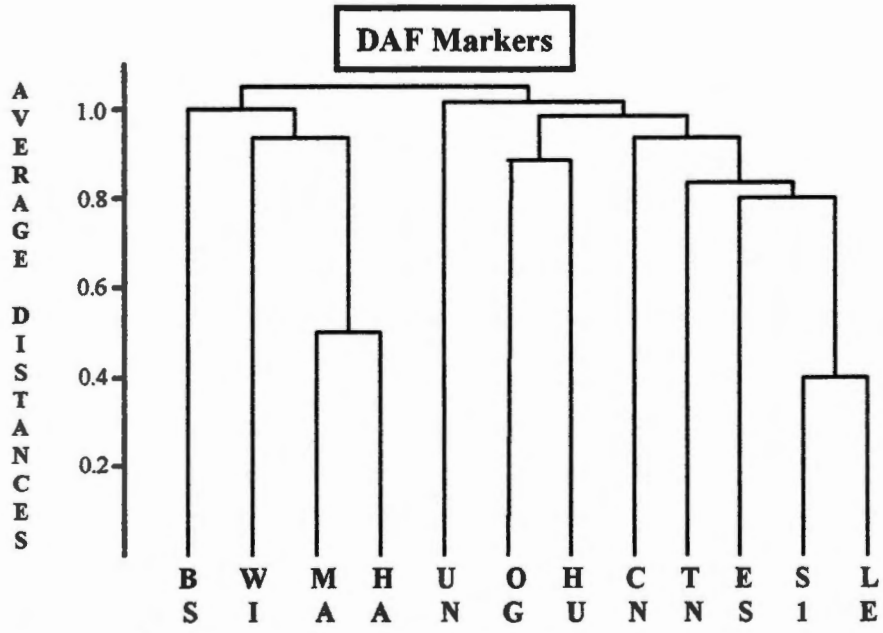


Figure 14. Clusters by DAF markers for northern and southern soybean cultivars or lines representing different eras. BS= BSR 101, WI= Williams, MA= Mandarin, HA= Harosoy, UN= Unknown, OG= Ogden, HU= Hutcheson, CN= CNS, TN= TN 5-85, ES= Essex, S1= S.100, LE= Lee

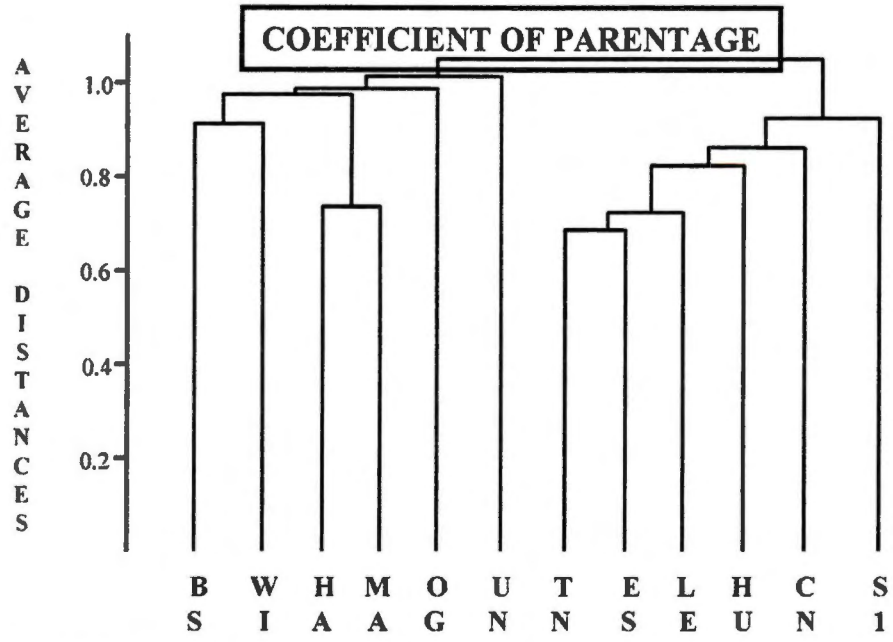


Figure 15. Clusters by coefficient of parentage for northern and southern soybean cultivars/lines representing different eras. BS= BSR 101, WI= Williams, MA= Mandarin, HA= Harosoy, UN= Unknown, OG= Ogden, HU= Hutcheson, CN= CNS, TN= TN 5-85, ES= Essex, S1= S.100, LE= Lee

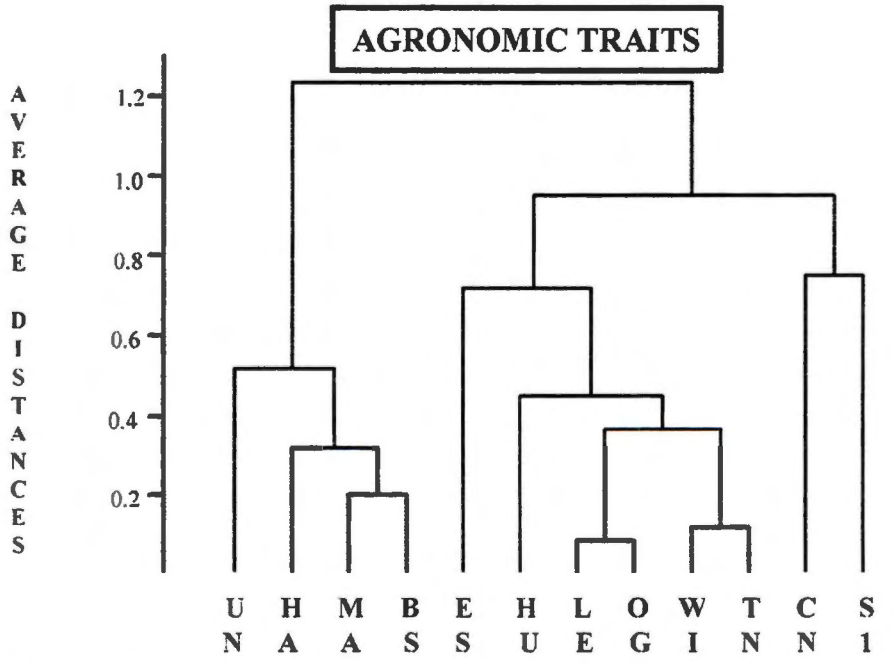


Figure 16. Clusters by agronomic traits for northern and southern soybean cultivars/lines representing different eras. BS= BSR 101, WI= Williams, MA= Mandarin, HA= Harosoy, UN= Unknown, OG= Ogden, HU= Hutcheson, CN= CNS, TN= TN 5-85, ES= Essex, S1= S.100, LE= Lee

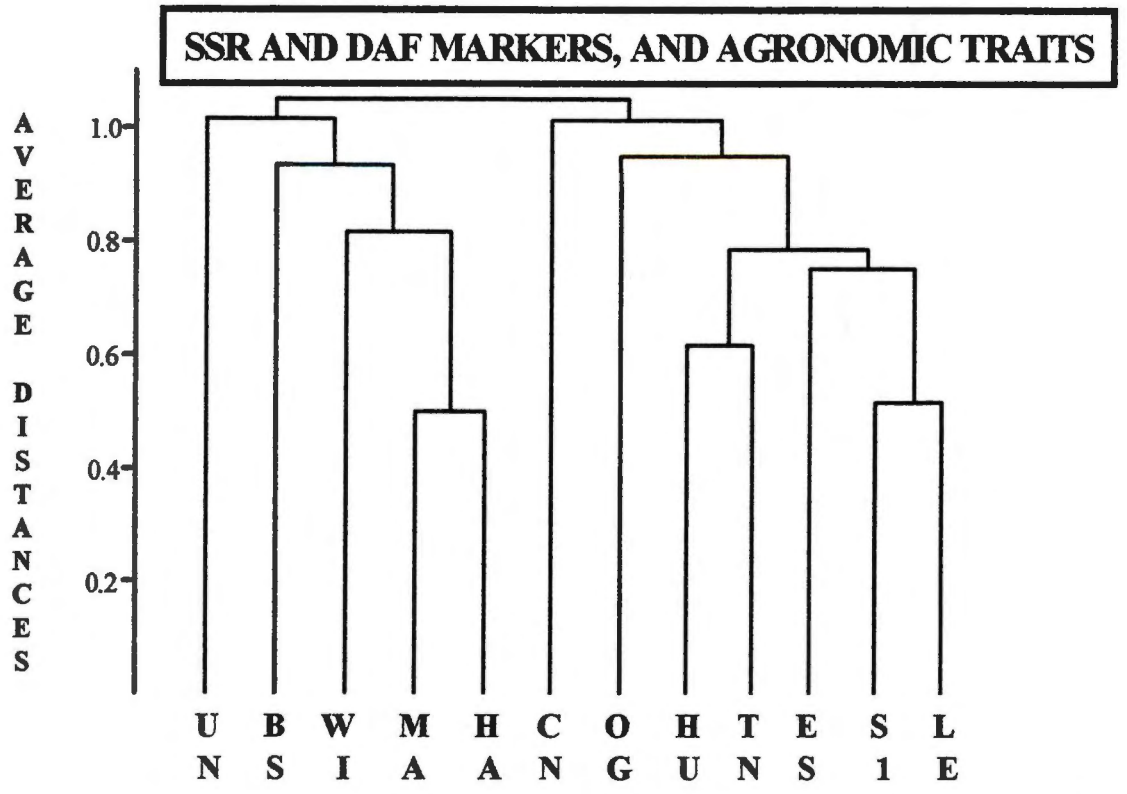


Figure 17. Clusters formed by the combination of DAF and SSR markers, and agronomic traits for northern and southern soybean cultivars or lines representing different eras. BS= BSR 101, WI= Williams, MA= Mandarin, HA= Harosoy, UN= Unknown, OG= Ogden, HU= Hutcheson, CN= CNS, TN= TN 5-85, ES= Essex, S1= S.100, LE= Lee

Table 12. Correlation coefficients among SSR-GS, DAF-GS, QT-GS and CP in soybean

	SSR-GS	DAF-GS	QT-GS	CP
SSR-GS	1.00	0.47**	0.57**	0.68**
DAF-GS		1.00	0.38**	0.63**
QT-GS			1.00	0.47**
CP				1.00

**correlation different from zero ($P \leq 0.01$)

CHAPTER 4

Conclusions

It is commonly known that soybean breeding in the U.S. started with a limited number of ancestral lines in the South and in the North. Six cultivars representing first, third and fifth generation cultivars were chosen for this study to represent different eras of development and release. Most of the cultivars have been used extensively as parents for successive generation of cultivars. Although there are some cases where plant introductions were used in crosses in order to get resistance to diseases, they did not cause a dramatic change in respect to genetic variability. In general, the results obtained in these experiments regarding broad sense heritabilities and molecular marker diversities imply that (1) there was not an apparent decline in genetic variability until the third generation cultivars (1970s), and (2) ancestral lines showed a considerable amount of variability.

Genetic variability in the southern soybean gene pool has declined, but this decline is not high. From ancestral lines to the fifth generation cultivars , the decline was 14 % for yield. Oil content, which had originally low variability, evidenced for 10 % decrease in genetic variation from ancestral lines to the fifth generation cultivars (Table 8).

SSR and DAF marker diversities diminished 58.3 % and 14.3 %, respectively. Molecular markers may not represent the desirable chromosome segments. Therefore, the decline indicated by molecular markers does not necessarily have practical importance for choosing the parents in crossing programs which can generate a high variation in segregating population for a given quantitative trait. On the other hand, the increase in genetic similarity may not cause severe reduction in genetic variance in segregating populations. As an example, the genetic variance for protein content between S.100 and CNS was zero (data not shown), but F_2 population of these lines showed a significant genetic variance. Selection in plant breeding has two genetic functions. First, to maintain the desired genes or chromosome

different marker types have been reported in soybean (Doldi *et al.* 1997, Prabhu *et al.* 1997 and Kisha *et al.* 1998), in wheat (Barret *et al.* 1998 and Burkhamer *et al.* 1998), and in corn (Smith *et al.* 1997, Senior *et al.* 1998, Pejic *et al.* 1998 and Ajmona Mason *et al.* 1998). However, Bohn *et al.* (1998) did not find any common pattern clusters formed by CP, RFLP, AFLP and SSR in wheat. Russell *et al.* (1997) assumed that SSR was not a suitable marker to assess genetic relationships in barley. The hexaploid wheat genome may give different results because each chromosome region has six copies. Some regions might be amplified much more than other regions. This may convey some deviations from expected results.

segments and secondly, to eliminate undesired genes. Elimination of undesired genes can cause decline in genetic variability for important traits due to linkage and pleiotropy. As a matter of fact, the genetic variability decline for yield in the southern gene pool does not seem a very serious problem for the current time.

Even though there was a decline in genetic variation for yield, apparently there appears to be considerable remaining variation to improve yield within southern germplasm. When protein and oil contents were considered the picture was different. These traits had relatively low heritabilities even at the beginning of soybean breeding. Furthermore, variation for oil content within southern germplasm revealed a declining trend.

If there was an efficient method to predict variation in segregating populations, cultivar improvement could be accelerated substantially by crossing only promising parents. Genetic distances/similarities among breeding lines by CP and QT-GS have limited efficiency for this purpose. Molecular markers have been thought to be a solution to this. However, previous research and this study did not prove it. SSR-GS correlated with heritability for yield. There might be two reasons for this relationship; (1) some of SSR markers may be associated with yield such as Satt079 (Mansur *at al.* 1996) and (2) small smaple size (n=8). Randomly chosen markers can be far away from the quantitative trait genes. As a result, they may not correlated with quantitative trait variation although they can cover the whole genome very well. Therefore, we suggest that changing the direction of studies from random markers to markers with QTL effects may give useful and better results.

Genetic similarities among the genotypes used in this study were estimated by SSR, DAF, CP and agronomic traits. Genetic similarity estimation by DNA markers was more efficient than CP and agronomic traits, but better comparison of CP with molecular markers can be done by running experiments with soybean cultivars released in the last 10-15 years. SSR markers were more effective than DAF markers to discriminate genotypes, but SSR markers most likely underestimate GS.

Molecular markers, CP and agronomic traits resulted in different clustering of southern and northern cultivars. Northern and southern soybean germplasm can be treated

as two different pools of germplasm. Cluster analysis, in general, classified cultivars/lines accordingly, although some disagreement occurred between clustering and GS estimations. There is a research need to determine the most efficient cluster method and related matters for genotype classification. When the raw molecular data and estimated GS were clustered, different groupings of genotypes came out. Another approach might be to develop software or SAS applications which can use discriminant analysis for molecular marker data.

Genetic similarity estimates by SSR, DAF, CP and agronomic traits showed significant correlations. When the cost of running DNA markers and biases in CP due to assumptions such as no relation between individuals with unknown pedigrees, it seems that the appropriate suggestion is to improve CP estimation with the help of molecular markers by changing CP assumptions. Using DAF or AFLP , soybean lines can be assayed and the GS can be used to modify coefficient of parentage between lines in pedigree with unknown relationships. As a result, zero coefficient of parentage may not be a necessary assumption anymore.

LITERATURE CITED

Ajmone-Marsan, P., P. Castiglioni, F. Fusari, M. Kuiper, and M. Motto. 1998. Genetic diversity and its relationship to hybrid performance in maize as revealed by RFLP and AFLP markers. Theor. Appl. Genet. 96:219-227

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

Allen, F. L. 1993.Usefulness of plant genome mapping to plant breeding. *In* P. M. Gresshoff (ed.) Plant genome analysis. CRC p. 11-18

Allen, F. L. and H. L. Bhardwaj. 1987. Genetic relationships and selected pedigree diagrams of North American soybean cultivars. The Univ. of Tennessee, Agricultural Experiment Station. Bulletin 652. Knoxville, TN

Barrett, B. A., K. K. Kidwell, and P. N. Fox. 1998. Comparison of AFLP and pedigree-based genetic diversity assessment methods using wheat cultivars from the Pacific Northwest. Crop Sci. 38:1271-1278

Bassam, B. J., G. Caetano-Anollés, and P. M. Gresshoff. 1991. Fast and sensitive silver staining of DNA in polyacrylamide gels. Anal. Biochem. 196:80-83

Bhatt, G. M. 1973. Comparison of various methods of selecting parents for hybridization in common bread wheat (*Triticum astivum* L.). Aust. J. Agric. Res. 24:457-464

Bohn, M., H. F. Utz, and A. E. Melchinger. 1999. Genetic similarity among winter wheat cultivars determined on the basis of RFLPs, AFLPs, and SSRs and their use for predicting progeny variance. Crop Sci. 39:228-237

Botstein, D., R. L. White, M. Skolnick, and R. W. Davis. 1980. Construction of a genetic map in man using restriction fragment length polymorphism. Am. J. Hum. Genet. 32:314-331

Burkhamer, R. L., S. P. Lanning, R. J. Martin, and L. E. Talbert. 1998. Predicting progeny variance from parental divergence in hard red spring wheat. Crop Sci. 38:243-248

Burton, J. W. 1987. Quantitative genetics : Results relevant to soybean breeding. *In* J. R. Wilcox (ed.), Soybeans : Improvement, production and uses. 2nd edition. Agron. Monogr. ASA, CSSA and SSSA, Madison, WI. p. 211-248

Caetano-Anollés, G., B. J. Bassam, and P. M. Gresshoff. 1991. DNA amplification fingerprinting using very short arbitrary oligonucleotide primers. Bio/Technol. 9:553-556

Carter Jr, T. E., Z. Gizlice, and J. W. Burton. 1993. Coefficient of parentage and genetic similarity estimates for 258 North American soybean cultivars released by public agencies during 1945-88. USDA Tech. Bull. No. 1814

Cox, T. S., Y. T. Kiang, M. B. Gorman, and D. M. Rodgers. 1985. Relationship between coefficient of parentage and genetic similarity indices in the soybean. Crop Sci. 25:529-532

Delaney, X., D. M. Rodgers, and R. G. Palmer. 1983. Relative genetic contributions among ancestral lines to North American soybean cultivars. Crop Sci. 23:944-949

Dellaporta, S. L., J. Wood, and J. B. Hicks. 1983. A plant DNA minipreparation. Version II. Plant Mol Biol. Rep. 1:19-21

Doldi, M. L., J. Vollman, and T. Lelley. 1997. Genetic diversity in soybean as determined by RAPD and microsatellite analysis. Plant Breeding 116:331-335

Falconer, D. S. and T. F. C. Mackay. 1995. Introduction to quantitative genetics. 4th ed. Longman Group, Ltd., Essex, England

Gizlice, Z., T. E. Carter, Jr, and J. W. Burton. 1993a. Genetic diversity in North American soybean: I - Multivariate analysis of founding stock and relation to coefficient of parentage. Crop Sci. 33:614-620

Gizlice, Z., T. E. Carter, Jr, and J. W. Burton. 1993b. Genetic diversity in North American soybean: II - Prediction of heterosis in F_2 populations of southern founding stock using genetic similarity measures. Crop Sci. 33:620-626

Gizlice, Z., T. E. Carter, Jr, and J. W. Burton. 1994. Genetic base for North American public soybean cultivars released between 1947 and 1988. Crop Sci. 34:1143-1151

Helms, T. C., J. H. Orf, and P. McClean. 1995. Genetic variance estimates in six soybean populations and genetic diversity measures. Agron. Abstracts. P.70. ASA, Madison, WI

Kisha, T. J., C. H. Sneller, and B.W. Diers. 1997. Relationship between genetic distance among parents and genetic variance in populations of soybean. Crop Sci. 37:1317-1325

Kisha, T. J., B. W. Diers, J. M. Hoyt, and C. H. Sneller. 1998. Genetic diversity among soybean plant introductions and North American germplasm. Crop Sci. 38:1669-1680

Luedders, V. D. 1977. Genetic improvement in yield of soybeans. Crop Sci. 17 : 971-972

Lorenzen, L. L., S. Boutin, N. Young, J. E. Specht, and R. C. Shoemaker. 1995. Soybean pedigree analysis using map based molecular markers: I- Tracking RFLP markers in cultivars. Crop Sci. 35:1326-1336

Manjarrez-Sandoval, P., T. E. Carter Jr., D. M. Webb, and J. W. Burton. 1997. Heterosis in soybean and its prediction by genetic similarity measures. Crop Sci. 37:1443-1452

Mansur, L. M., J. H. Orf, K. Chase, T. Jarvik, P. B. Cregan, and K. G. Lark. 1996. Genetic mapping of agronomic traits using recombinant inbred lines of soybean. Crop Sci. 36:1327-1336

Milligan, G.W. and Cooper, M.C. 1987. A study of variable standardization. College of Administrative Science Working Paper Series, 87-63, The Ohio State University, Columbus, OH.

Nei, M., and W. H. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. (USA) 76:5269-5273

Pejic, I., P. Ajmone-Marsan, M. Morgante, V.Kozumplick, P. castiglioni, G. Taramino,and M. Motto. 1998. Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs, and AFLPs. Theor. Appl. Genet. 97:1248-1255

Prabhu, R. R., D. Webb, H. Jessen, S. Luk, S. Smith, and P. M. Gresshoff. 1997. Genetic relatedness among soybean phenotypes using DNA amplification fingerprinting (DAF), RFLP and pedigrees. Crop Sci. 37:1590-1595

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

Russell, J. R., J. D. Fuller, M. Macaulay, B. G. Hartz, A. Jahoor, W. Powell, and R. Waugh. 1997. Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. Theor. Appl. Genet. 95:714-722

Senior, M. L., J. P. Murphy, M. M. Goodman, and C. W. Stuber. 1998. Utility of SSRs for determining genetic similarities and relationships in maize using an agorose gel system. Crop Sci. 38:1088-1098

Shoemaker, R. C., R. D. Guffy, L. L. Lorenzen, and J. E. Specht. 1992. Molecular genetic mapping of soybean: Map utilization. Crop Sci. 32:1091-1098

Shoemaker, R. C., L. L. Lorenzen, B. W. Diers, and T. C. Olson. 1993. Genome mapping and agriculture. *In* P. M. Gresshoff (ed.). Plant genome analysis. CRC Press. p. 1-10

Skorupska, H. T., R. C. Shoemaker, A. Warner, E. R. Shipe, and W. C. Bridges. 1993. Restriction fragment length polymorphism in soybean of the southern USA. Crop Sci. 33:1169-1176

Smith, J. S. C., E. C. L. Chin, H. Shu, O. S. Smith, S. J. Wall, M. L. Senior, S. E. Mitchell, S. Kresovitch, and J. Ziegle. 1997. An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.) : Comparisons with data from RFLPs and pedigree. Theor. Appl. Genet. 95:163-173

Sneller, C. H., J. W. Miles, and J. M. Hoyt. 1997. Agronomic performance of soybean plant introductions and their genetic similarity to elite lines. Crop Sci. 37:1595-1600

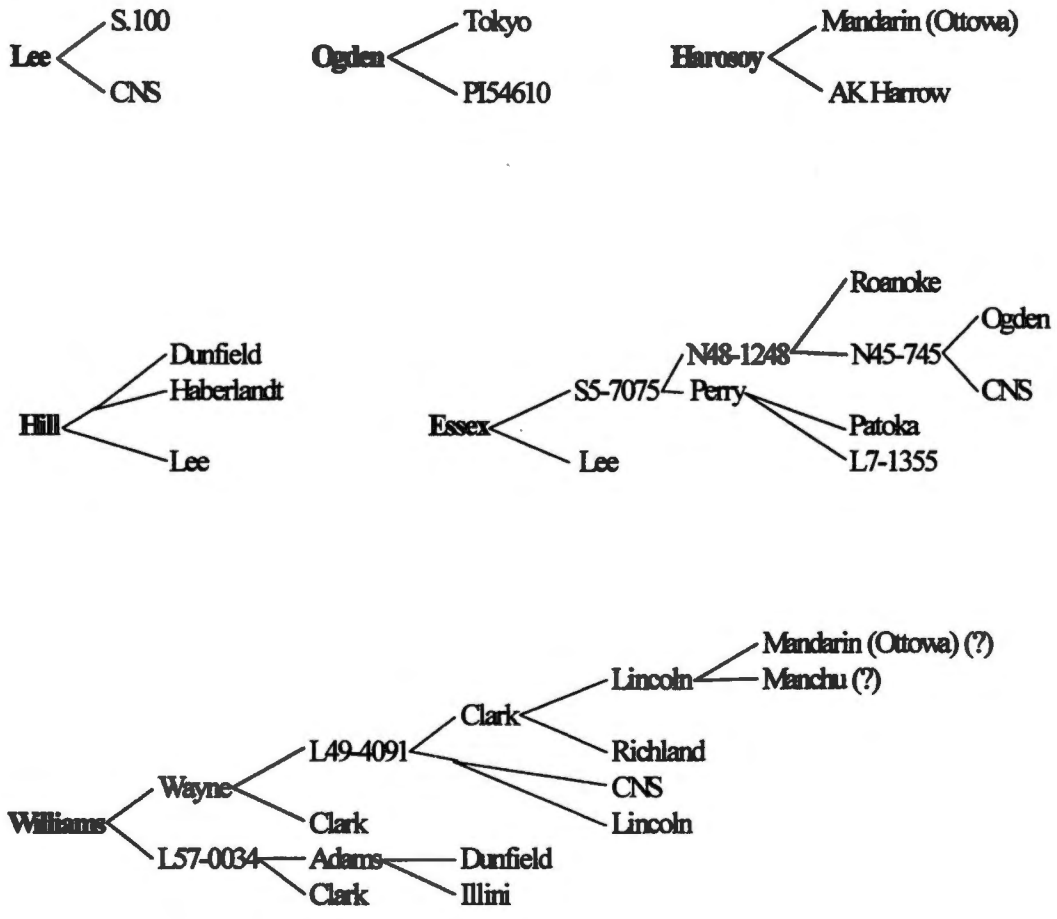
St. Martin, S. K. 1982. Effective population size for the soybean improvement program in maturity groups 00 to IV. Crop Sci. 22:151-152

Tautz, D. 1989. Hypervariability of simple sequences as a general source of polymorphic DNA markers. Nucl. Acids Res. 17:6463-6471

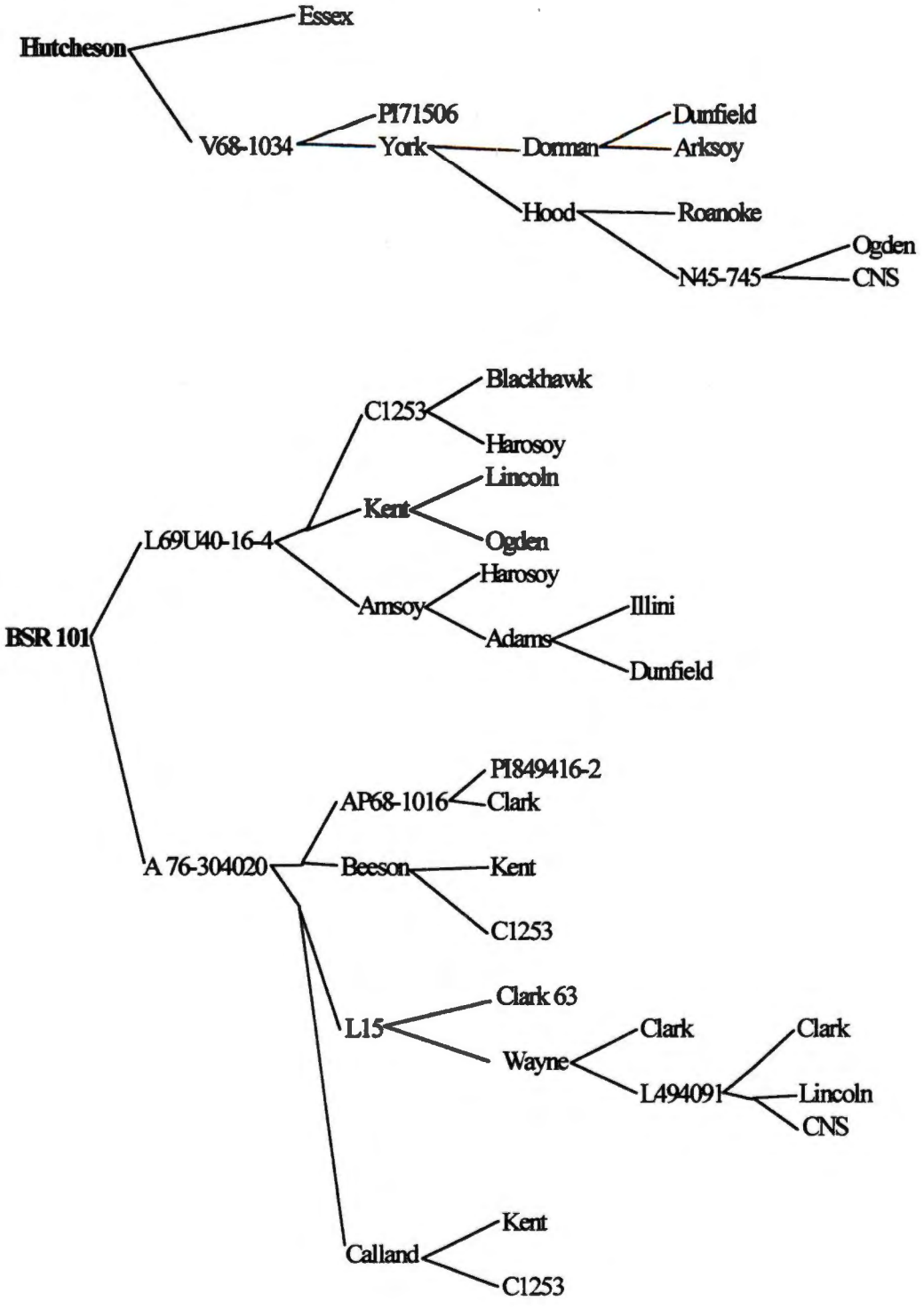
Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. Nuc. Acids Res. 23:4407-4414

Williams, J. G. K., A R. Kubelik, K. J. Lavak, J. A. Rafalski, and S. Y. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucl. Acids Res. 18:6531-6535

APPENDICES



Appendix 1. Pedigree diagrams of Lee, Ogden, Harosoy, Hill, Essex, Williams and TN 5-85



Appendix 2. Pedigree diagram of Hutcheson and BSR 101

Appendix 3. Genotypic variance (σ^2_g), phenotypic variance (σ^2_p) and broad sense heritability (h^2) for maturity and plant height in different F_2 populations of soybean

Population	σ^2_g	σ^2_p	h^2	Population	σ^2_g	σ^2_p	h^2
Maturity							
Southern x southern				**Northern x southern**			
S.100 x CNS	75.5	85.5	0.67	Mandarin x CNS	153.1	157.8	0.97
Lee x Ogden	69.7	79.8	0.87	Harosoy x Lee	116.0	121.0	0.96
Essex x Unknown	35.0	50.0	0.70	Williams x Essex	97.1	110.6	0.88
TN 5-85 x Hutcheson	57.1	64.8	0.88	BSR 101 x TN 5-85	183.8	188.8	0.97
Plant height							
Southern x southern				**Northern x southern**			
S.100 x CNS	553.6	651.0	0.85	Mandarin x CNS	514.9	576.2	0.85
Lee x Ogden	105.7	183.6	0.58	Harosoy x Lee	552.3	641.6	0.86
Essex x Unknown	80.9	123.7	0.65	Williams x Essex	272.9	322.0	0.85
TN 5-85 x Hutcheson	50.2	112.7	0.45	BSR 101 x TN 5-85	211.9	262.2	0.81

Appendix 4. Genetic similarities estimated by quantitative traits among soybean genotypes

	Cultivars											
	MA	HA	WI	BS	CN	S1	LE	OG	ES	UN	HU	TN
MA	1.000											
HA	0.982	1.000										
WI	0.618	0.759	1.000									
BS	0.985	0.997	0.741	1.000								
CN	0.000	0.232	0.812	0.183	1.000							
S1	0.691	0.820	0.973	0.798	0.763	1.000						
LE	0.285	0.482	0.944	0.456	0.942	0.889	1.000					
OG	0.297	0.490	0.942	0.462	0.952	0.889	0.998	1.000				
ES	0.451	0.616	0.976	0.600	0.878	0.925	0.983	0.982	1.000			
UN	0.866	0.940	0.902	0.939	0.480	0.936	0.718	0.713	0.815	1.000		
HU	0.450	0.618	0.976	0.606	0.842	0.928	0.975	0.966	0.991	0.834	1.000	
TN	0.484	0.650	0.986	0.634	0.852	0.949	0.976	0.969	0.991	0.851	0.998	1.000

MA= Mandarin, HA= Harosoy, WI= Williams, BS= BSR 101, CN= CNS, S1= S.100, LE= Lee, OG= Ogden, ES= Essex, UN= Unknown, HU= Hutcheson, TN= TN 5-85

Appendix 5. Coefficient of parentage among soybean genotypes

	Cultivars											
	MA	HA	WI	BS	CN	S1	LE	OG	ES	UN	HU	TN
MA	1.000											
HA	0.250	1.000										
WI	0.000	0.000	1.000									
BS	0.063	0.125	0.117	1.000								
CN	0.000	0.000	0.031	0.004	1.000							
S1	0.000	0.000	0.000	0.000	0.000	1.000						
LE	0.000	0.000	0.016	0.002	0.250	0.250	1.000					
OG	0.000	0.000	0.000	0.063	0.000	0.000	0.000	1.000				
UN	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000			
ES	0.000	0.000	0.010	0.005	0.156	0.125	0.266	0.031	0.000	1.000		
HU	0.000	0.000	0.010	0.007	0.094	0.063	0.141	0.031	0.000	0.270	1.000	
TN	0.000	0.000	0.011	0.008	0.141	0.125	0.258	0.047	0.000	0.326	0.173	1.000

MA= Mandarin, HA= Harosoy, WI= Williams, BS= BSR 101, CN= CNS, S1= S.100, LE= Lee, OG= Ogden, ES= Essex, UN= Unknown, HU= Hutcheson, TN= TN 5-85

PART III

GENETIC VARIATION IN F_2 AND F_3 SOYBEAN POPULATIONS IN NORTH AMERICA BY QTL ASSOCIATIONS WITH AGRONOMIC TRAITS

Abstract

Soybean is a relatively new crop in the U.S. compared to other crops. In the last 60-70 years, a considerable amount of resources has been invested in soybean breeding. As a result of this investment and soybean breeders' determination, genetic progress in yield and other traits have been achieved. It has been suggested that genetic variability has decreased in soybean in North America. U.S. soybean cultivars can be traced back to 50 ancestral lines at the most. Of the 50, 14 ancestral lines contributed approximately 80 % of the genes that are present in modern cultivars. The genetic background goes back to a limited number of parents. Therefore, this study was initiated to: (1) show the genetic trends by estimation of heritabilities in soybean populations created between southern x southern, and northern x southern cultivars/lines representing different eras, (2) find and compare microsatellite (SSR) QTLs (Quantitative Trait Locus) with agronomic traits in the cross between southern ancestral lines and between fifth generation southern cultivars, and (3) to compare QTLs found in F_2 with the ones found in F_3.

Eight F_2 and $F_{2:3}$ populations were developed by crossing ancestral lines and cultivars representing different eras. Two sets of four crosses were made between southern x southern, and between northern x southern cultivars/lines. F_2 plants and $F_{2:3}$ lines were grown at Knoxville Experiment Station in 1997 and 1998, respectively. Maturity, plant height, protein and oil content, and yield were recorded on F_2 plants, $F_{2:3}$ lines and parents of the crosses.

Heritabilities for yield estimated by parent-offspring regression revealed a low variation for southern x southern crosses. Therefore, it was not possible to observe a genetic trend. Variability for protein and oil content , and plant height was either increased or stayed on the same level within southern soybean breeding lines compared to the ancestral lines. A small decrease was observed in genetic variability for maturity. Northern x northern populations exhibited reasonably high heritability for yield, but their population means were

lower than southern x southern crosses. Other traits measured in northern x southern crosses did not reveal any pattern for cultivar generations. The cross between Ogden and Lee indicated exceptionally high heritabilities for yield as well as protein and oil content. It also had the highest population mean for yield among all crosses in F_2 and F_3 generations.

The cross between the two southern ancestral lines and the cross between fifth generation southern cultivars were compared for genetic variability by QTL associations of four SSR markers with agronomic traits. In each population, three QTLs were found. However the QTLs in crosses were different. The only common QTL was the association of Satt022 with yield. This QTL showed evidence for dominant effects in both populations. By considering the overall results, it is concluded that the detection of genetic variability may be more consistent by QTL analysis than by the heritability estimation.

CHAPTER 1

Introduction and Literature Review

The objectives of soybean breeding programs can be expressed in both biological and economical terms. Biological objectives are dominated by yield, oil and protein content, and resistance to biotic stresses. Achievements in biological objectives can at the same time serve the economic objectives. The main steps in fulfilling the objectives satisfactorily in plant breeding are : (1) choosing suitable parents, (2) exploiting variation, (3) making selections, and (4) determining the superior genotypes.

Decision on parents for crossing and exploiting the variation depend upon the degree of variation for qualitative and quantitative traits within the breeding lines. Plant breeding alters the genetic composition of crops within the short term compared to natural selection. This alteration can cause decline in genetic variability. It is important to know how much variation remains after repeated selections over time and what is the rate of decline. The knowledge can be very helpful to make genetic progress for yield and other traits and to make projections for the future.

Soybean has become one of the five important crops in the U.S. in the last 70 years mostly due to genetic improvements. This genetic improvement has been achieved within a narrow genetic background(St Martin, 1982, Delanay *et al.* 1983 and Gizlice *at al.* 1993a). Modern soybean cultivars can be traced back to a limited number of ancestral lines. In addition, crossing of similar lines and selection within the same genetic background caused the decline in soybean genetic variability for agronomic traits (St Martin, 1982, Delanay *et al.* 1983, Allen and Bhardwaj, 1987, and Gizlice *et al.* 1993a,b).

Heritability estimation, the ratio of genetic variance to phenotypic variance, is specific to the reference population and it relates to the particular set of environmental conditions. That makes difficult to generalize the result for whole genotypes within a crop (Dudley and

Moll, 1969). However, it is the only tool to predict selection progress in breeding programs. Burton (1987) reviewed the studies on heritabilities of soybean traits. Heritability for yield ranged from 0.03 to 0.58 in different studies. Heritabilities for protein content, oil content, plant height and maturity were between 0.51 and 0.94.

Research on plant molecular genetics has yielded new opportunities or tools to plant breeders and other scientists which play roles in technology generation in agriculture. One of the significant contributions of plant geneticists to applied science was to make available molecular markers. As a quite new tool, molecular markers are expected to expedite plant breeding progress. Griffin and Palmer (1987) stated four characteristics for an ideal genetic marker: (1) the alleles should be completely penetrated, (2) the genetic expression of the marker should be codominant, (3) marker loci should not show any epistatic interaction, and (4) marker genotypes should not have any deleterious effect on plant vigor. In addition to these characteristics, a sufficient number of polymorphic loci and reliability of markers should be considered very carefully. The second characteristic also relates to the required sample size. Carlson *et al.* (1991) calculated that 280 samples are needed to detect an identical linkage in dominant markers; whereas 50 samples are sufficient for the same purpose in codominant markers.

Among available DNA markers, Restriction Fragment Length Polymorphism (RFLP; Bolstein, 1980) and mirosatellite (SSR; Tautz, 1988) markers mostly show codominant inheritance. SSR markers exhibit more polymorphisms than RFLP and they are cheaper and easier to test than RFLP. A small amount of DNA requirement and suitability for automation are the other advantages of SSR markers over RFLP. SSR markers were proven to be very repeatable; whereas a dominant marker, RAPD (Random Amplified Polymorphic DNA) produced different results from different laboratories (Jones *et al.* 1997).

Molecular markers have been employed in plant genetics for different purposes such as to determine the associations between quantitative traits (QTL; Quantitative Trait Loci), to establish genetic maps, marker assisted selection and so on. Quantitative traits are

controlled by many alleles. Thus, there is no clear-cut distinction among phenotypes. Their complex genetic structure makes visual selection difficult and slows the gain from selection. The idea behind linkages between molecular markers and major genes of quantitative traits has attracted plant breeders to carry out studies to find, locate and map QTLs. Specific linkages to breeding populations is one of the constraints in molecular marker studies. Thus, Skorupska *et al.* (1993) proposed the molecular marker evaluation of a wide range of soybean genotypes. It is also important to conduct experiments with cultivated genotypes which show relatively low polymorphisms compared to interspecific crosses. However, research on wild genotypes or populations derived from cultivated x wild species might provide insights into the understanding of quantitative trait inheritance.

Sax (1923) reported the association between a quantitative trait (seed size) and a monogenic one (seed coat pigmentation) in beans. He gave the cause of this correlation as linkage of polygenic characters to a particular locus. A number of studies have been carried out on this matter, but weaknesses in morphologic markers have resulted in limited success. The advent of molecular markers led to the increase in the number and magnitude of QTL research. The main idea behind the QTL research in crops is to make marker assisted selection feasible in plant breeding as well as to enlighten some phenomenon related to quantitative traits such as heterosis and epistasis. A number of studies have been conducted on QTL in soybeans (Keim *et al.* 1990, Diers *et al.*1992, Mansur *et al.* 1993, Lark *et al.* 1994, Lee *et al.*1996a,b, Mansur *et al.* 1996, Bailey *et al.* 1997, Brummer *et al.* 1997, Mian *et al.* 1998, and Lewers *et al.* 1999). Soybean QTLs for different traits were found and mapped mostly by RFLP markers.

Studies on the inheritance of quantitative traits have been conducted to estimate genetic variances and heritabilities in soybean. Genetic variances for different traits have been estimated to help breeders in selection. There are no studies which show how the genetic variances and heritabilities for yield, protein and oil have changed over time in soybean. Although there have been suggestions and assumptions based on the selection theory that

additive variance would decrease over time and parallel to this, gain from selection would be reduced, New and better crop cultivars emerges from the breeding programs. Creating crosses to represent different eras in soybean breeding history could serve to indicate the trend in genetic variation for quantitative traits in soybean.

QTL studies have supplied very useful information by showing the location of some major genes in different crops including soybean. Keim *et al.* (1990) associated 20 RFLP markers with eight traits in soybean. Diers *et al.* (1992) determined the locations of eight and nine soybean QTLs for protein and oil, respectively. Mansur at al. (1993) showed three marker loci correlated to the yield, maturity and plant height. Lark *et al.* (1994) found a RFLP locus associated with oil and protein. Lee *at al.* (1996b) identified 13 QTLs for protein in five linkage groups and six QTLs for oil in three linkage groups. Mian *at al.* (1996) found 16 RFLP markers associated with seed weight. Brummer *at al.* (1997) located 11 and 13 QTLs by RFLP markers for oil and protein, respectively. Eleven QTLs for oil dispersed on five linkage groups and 13 QTLs for protein were on nine linkage groups. Accumulated information on QTL at the moment is not sufficient to use for practical purposes such as marker assisted selection for polygenic traits like yield, protein content and oil content.

QTL research can supply useful information to breeders by demonstrating the conserved regions of soybean genome over long term selection. Also, QTL can be used to reveal the information on genetic loss or gain for quantitative traits in crops. The result of selection in plant breeding is to accumulate desired genes for agronomic traits and eliminate the genes for undesired traits such as lodging and disease susceptibility. However, it is not a clear cut process due to linkages between the genes on the same chromosome. In the elimination process, some desired genes can also be eliminated. This genetic loss may narrow the genetic variability. On the other hand, selection for desired traits can increase the frequency of desired genes on the genome. This genetic gain may result in better progenies from better parents. Comparison of QTLs in soybean breeding lines from different decades may verify the genetic gain and loss or genetic shift of agronomic traits.

The objectives of this study were: (1) to show the genetic variability trends in soybean germplasm by estimation of additive genetic variances and heritabilities in soybean populations created between southern x southern, and northern x southern cultivars/lines representing different cultivar generations, (2) to find and compare QTLs by SSR markers for agronomic traits in two populations generated by crossing two southern ancestral lines, as well as the fifth generation cultivars, and (3) to compare QTLs found in F_2 plants with QTLs in $F_{2:3}$ lines.

CHAPTER 2

Materials and Methods

Cultivars/Lines and Crosses

Southern and northern ancestral lines and cultivars, CNS, S.100, Ogden, Lee, Unknown (Hill), Essex, TN 5-85, Hutcheson, Mandarin (Ottawa), Harosoy, Williams, and BSR 101, were chosen for inclusion in this study (Table 13). The seeds were obtained from Dr. Randal L. Nelson, soybean curator, USDA, ARS, Urbana, Illinois. The twelve entries

Table: 13- Selected southern and northern cultivars/lines, and their maturity group (MG), cultivar generation, release year and pedigree.

Cultivar/Line	Abbreviation	MG	Cultivar Generation	Release Year	Pedigree‡
Southern					
CNS	CN	VII	Ancestor	-	Introduction
S.100	S1	V	Ancestor	-	Introduction
Ogden	OG	VI	1	1953	Tokyo x PI 54610
Lee	LE	VI	1	1954	S.100 x CNS
Hill (Unknown)	UN	-	-	-	-
Essex	ES	V	3	1972	Lee x S5-7075
TN 5-85	TN	V	5	1986	Essex x D68-127
Hutcheson	HU	V	5	1987	Essex x V68-1034
Northern					
Mandarin+	MA	0	Ancestor	-	Introduction
Harosoy	HA	II	1	1951	Mandarin+ x AK Harrow
Williams	WI	III	3	1971	Wayne x L57-0034
BSR 101	BS	I	5	1988	A76-304020 x L69U40-16-4

+Mandarin (Ottawa)

‡ The complete pedigrees tracing back to ancestral lines are presented in Appendix 1 and 2.

represent ancestral lines, and selected first, third and fifth generation southern and northern cultivars representing different eras of production from the 1950's to the 1990's. The lines/cultivars have been chosen based on three criteria; The ancestral lines CNS, S.100 and Mandarin were chosen because they are major contributing lines to the modern gene pool of North American soybean cultivars (Delanay *at al.* 1983 and Gizlice *at al.* 1993a). The cultivars were chosen because they represent different generations of related cultivars and they were widely grown during their era of production. In addition, most of them have been utilized extensively as parents in both public and private breeding programs.

Crosses were made between the two southern ancestral lines between one southern and one northern ancestral lines (Table 14). Crosses were also made between first, third and fifth generation southern x southern or northern x southern cultivars (Table 14). Twenty F_1 seeds were sent to Costa Rica to produce F_2 seeds in winter nurseries. In addition, 50 seed from each F_1 plant were space-planted at Knoxville in 1997 in 3 m long rows in an incomplete block design. The plant spacing within rows was approximately 6 cm with 76 cm between rows. Approximately 250 plants from each F_2 population and 50 plants from each parent were selected randomly at R1 stage and tagged. Plant height, maturity, protein and oil content, and yield were recorded on selected plants.

Seeds from selected F_2 plants were planted to grow $F_{2:3}$ lines at Knoxville in 1998 in 3m long rows in an incomplete block design. $F_{2:3}$ lines from each populations were equally represented in each block and 12 parents was randomized with the lines in each block. The parents and $F_{2:3}$ lines in each block were randomized within the block by SAS procedure SORT (SAS Institute, Inc. 1998, Cary, NC). Parent lines were placed in each block to estimate the experimental error as accurately as possible. The spacing within the row and between the rows was approximately 4 cm and 76 cm, respectively. At harvest, 50 cm from each end of rows was discarded to eliminate border effects. The traits recorded on F_2 plants were also recorded in each $F_{2:3}$ row.

Table 14. The list of crosses made between southern x southern and northern x southern soybean cultivars/lines

Cross	Cultivar generation	Type of cross
S.100 x CNS	Ancestral lines	Southern x southern
Mandarin x CNS	Ancestral lines	Northern x southern
Lee x Ogden	First	Southern x southern
Harosoy x Lee	First	Northern x southern
Essex x Unknown	Third	Southern x southern
Williams x Essex	Third	Northern x southern
TN 5-85 x Hutcheson	Fifth	Southern x southern
BSR 101 x TN 5-85	Fifth	Northern x southern

DNA Analysis

Very young leaves of the selected F_2 plants were collected at R1 stage. One hundred samples were chosen randomly from S.100 x CNS and TN 5-85 x Hutcheson populations. Genomic DNA was isolated according to Dellaporta *et al* (1983). DNA concentration was measured in a TK0100 Dedicated Mini-Flourometer (Hoeffer Scientific Instruments, San Francisco, CA) by following manufacturer instructions.

Sixty five SSR primers were tested for polymorphisms in both populations. Only five primer sets showed segregation in both populations. SSR loci (primers) were Satt022, Satt050, Satt180, Satt236 and Satt353. They were placed in N, A1, C1, A1 and H linkage groups in USDA/Iowa State University Map, respectively. PCR reaction mixes were adjusted according to protocol from Research Genetics (Huntsville, AL) and primer pairs were synthesized by Research Genetics (Huntsville, AL). Ten µl reaction mixes were employed for each sample. Ten µl reaction mixes included 1 µL 10x PCR buffer, 1.6 µL dNTPs (1.25 mM), 0.9 µL $MgCl_2$ (25 mM), 0.3 µL forward primer, 0.3 µL reverse primer, 0.06 µL Taq DNA

polymerase (5 Units μL^{-1}), 2 μL DNA (10 ng μL^{-1}) and 3.84 μL sterile di H_2O. The amount of Taq DNA polymerase was 0.045 μL in the original protocol. We increased it to 0.06 μL to get better DNA amplification. PCR protocol was adjusted according to Rongwen *et al.* (1995). DNA was amplified in a MJ Research PTC-200 DNA Engine (MJ Research, Watertown, MA). PCR products were separated on a 5% polyacrylamide gel containing 7 M urea. Detection and visualization of markers were fulfilled as described by Bassam *at al.* (1991), but gels were run for 3.5-4 h at 350 W using a Protean gel chamber (Bio-Rad Inc., Richmond, CA). In silver staining, fixation and staining stages were increased to 15 and 30 min, respectively. After gels were dried, homozygous dominant, heterozygous and homozygous recessive individuals were denoted as one, two and three, respectively.

Statistical Analysis

Variance components in the mixed model were estimated. However, variance component estimation was biased in this study because the parent data from 1997 and 1998 were not similar. Even the logarithmic transformation did not solve the problem particularly for yield. Therefore, parent-offspring regression was applied to estimate narrow sense heritability in SAS PROC REG. When offsprings ($F_{2:3}$ lines) are regressed on the parents (F_2 plants), the regression coefficient will equal half the narrow sense heritability (Falconer and Mackay, 1995). Parent-offspring regression is expected to give less biased estimation of narrow sense heritability when the F_2 and F_3 generation grown in different years because the environmental variance will be deduced only from parent data.

Yield, protein content, oil content, maturity and plant height recorded from F_2 plants were tested in each population against SSR marker classes by one way ANOVA in SAS procedure GLM. Additive and dominance effects were estimated by the contrast statement (Liu, 1998). The model for QTL was $y=\mu + g + e$ where y, μ, g and e are quantitative trait, general mean, marker genotypes, and experimental error, respectively. SSR marker data in ANOVA model can be coded in different ways either as 1 being homozygous dominant (AA),

2 heterozygous (Aa) and 3 homozygous recessive (aa). The additive and dominant effects were tested by the SAS contrast statement. In the statement, the linear functions for additive and dominant effects were 1 0 -1 and -1 2 -1, respectively. The decision on the significance for the F_2 data was made based on the additive and dominant effects. The data for the same traits from $F_{2:3}$ rows were also computed to determine QTLs. However, the data from heterozygous individuals in F_2 were discarded in QTL analysis for $F_{2:3}$, because they would segregate in $F_{2:3}$ generation. Therefore, there were only two marker classes, homozygous dominant and homozygous recessive.

CHAPTER 2

Results and Discussion

Heritability estimates

Hybridization in plant breeding aims to create high genetic variation with high population mean. Southern x southern crosses compared to the northern x southern crosses showed more range in yield and higher means (Figure 18 and 19). Compared to the ancestral line population, the first and the fifth generation population had a higher population means. In northern x southern crosses, the first and third generation crosses exhibited significantly higher yield means than ancestral line crosses.

Narrow sense heritabilities reflect the additive genetic variance in segregating populations. Therefore, they are very useful to estimate the gain from a selection for a given trait. There are several approaches to estimate additive variance and heritabilities. Of the estimation procedures, regression of offspring on parents is straight forward. In this experiment, F_2 plants and $F_{2:3}$ lines represent parents and offsprings, respectively. The heritabilities for yield was not significant with the exception of the Lee x Ogden population in the southern x southern crosses. In the third and fifth generation crosses, there was not any genetic additive variation (Table 15). All of the northern x southern crosses exhibited significant heritabilities for yield. The heritabilities for yield in southern x southern and northern x southern crosses did not show any trend on cultivar generations. In other words, the genetic variability did not show decline or increase over time.

The absence or non-significance of additive genetic variance in southern x southern crosses can not be explained by the presence of very high dominance variance, because in that case, there would have been more research on hybrid soybean cultivars to make it feasible. Northern x southern crosses would be expected to show higher means than southern x southern crosses. The second possibility for this could be high experimental error that

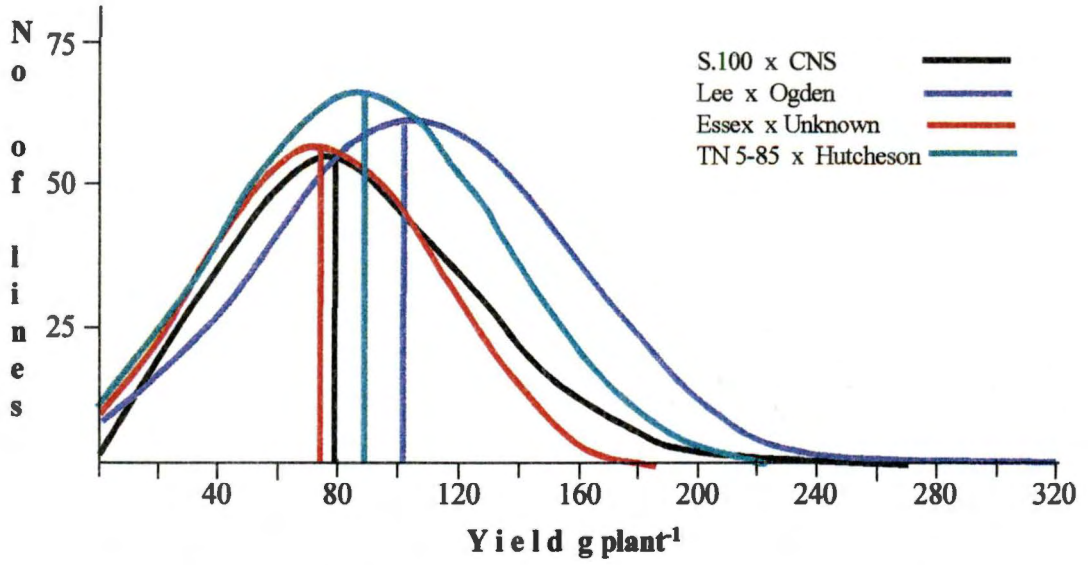


Figure 6. The distribution of yield and population mean in southern x southern soybean F_2 populations

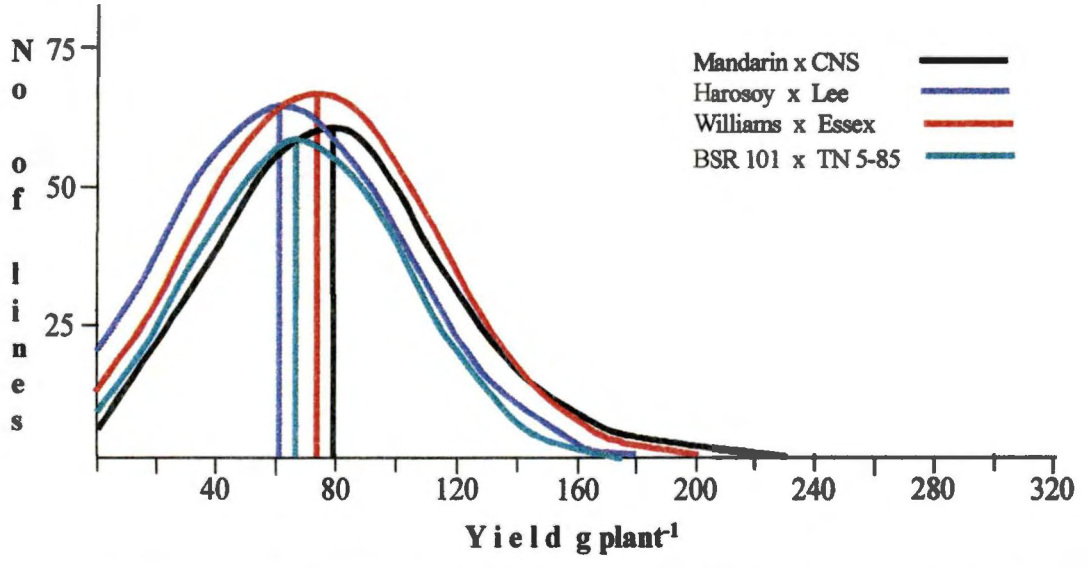


Figure 7. The distribution of yield and population mean in northern x southern soybean F_2 populations

Table 15. Narrow sense heritabilities (h^2) and standard errors (Se) for heritabilities by parent-offspring regression for seed yield, protein content and oil content in different F_2 and $F_{2:3}$ soybean populations representing different eras in soybean breeding.

Population	$h^{2(1)}$	Se	Population	h^2	Se
Southern x Southern			**Northern x Southern**		
			Yield		
S.100 x CNS	0.11	0.12	Mandarin x CNS	0.31**	0.09
Lee x Ogden	0.26**	0.09	Harosoy x Lee	0.37**	0.09
Essex x Unknown	0.00	0.10	Williams x Essex	0.39**	0.09
TN 5-85 x Hutcheson	0.00	0.09	BSR 101 x TN 5-85	0.37**	0.13
			Protein content		
S.100 x CNS	0.22	0.15	Mandarin x CNS	0.63**	0.25
Lee x Ogden	1.36**	0.18	Harosoy x Lee	0.00	0.00
Essex x Unknown	0.38**	0.18	Williams x Essex	0.33	0.34
TN 5-85 x Hutcheson	0.60**	0.17	BSR 101 x TN 5-85	0.33*	0.13
			Oil content		
S.100 x CNS	0.16	0.16	Mandarin x CNS	0.99**	0.13
Lee x Ogden	1.21**	0.16	Harosoy x Lee	0.00	0.00
Essex x Unknown	0.72**	0.23	Williams x Essex	1.27**	0.43
TN 5-85 x Hutcheson	0.62**	0.20	BSR 101 x TN 5-85	0.57*	0.29

*, ** significant ($P \leq 0.05$ and $P \leq 0.01$, respectively)

(1) $h^2 = 2 \times b$, where b is the regression coefficient.

obscures the genetic differences between the lines. However, the experimental error was not the reason for the non-significant heritabilities. In parent-offspring regression $b = 1/2\ h^2$, $b = COV_{OP} / V_P$ and $COV_{OP} = 1/2\ V_A$ where b= regression coefficient, h^2= narrow sense heritabilty, COV_{OP}= covariance between parent and offspring, V_P= variance of parents, and V_A=additive variance (Falconer and Mackay, 1995). V_P was deduced only from F_2 plants. The broad sense heritabilities for yield in F_2 were very high (Table 6). The genetic variation was much better exploited in F_2 in 1997, but it was not exploited in $F_{2:3}$ lines in 1998 because of environmental conditions.

The protein content and oil content within southern soybean gene pool was increased through the fifth cultivar generation compared to ancestral lines. The cross between two major ancestors, S.100 and CNS, showed a non-significant heritability for protein content and oil content (Table 15). Lee x Ogden population showed very high heritability for protein and oil content (1.36 and 1.21, respectively). It is possible to get heritabilities higher than one in parent-offspring regression. In northern x southern crosses, Harosoy x Lee was very poor in genetic variability for protein and oil content. Williams x Essex population had exceptionally high heritability for oil content but a non-significant one for protein content. The results did not show either decreasing or increasing trend over time for protein content and oil content in northern x southern populations.

Maturity and plant height normally have high heritabilities in soybean (Burton, 1987). All the heritabilities for these two traits were significant but not in S.100 x CNS population for plant height (Appendix 6). Heritability range was from 0.39 (TN 5-85 x Hutcheson) to 1.47 (Harosoy x Lee) for maturity and from 0.05 (S.100 x CNS) to 1.09 (BSR 101 x TN 5-85). The variation still remains in the fifth generation cultivars. The genetic variability seemed to be decreased in the fifth generation cross, TN 5-85 x Hutcheson, for maturity but it was still high and significant.

QTL Analysis

Molecular markers are assumed to be free of environmental effects, but agronomic traits are affected by the environmental conditions considerably. Therefore, the associations (QTL) between molecular markers and agronomic traits are also affected by the environment. In S.100 x CNS F_2 plants, there were two significant QTLs (Satt022 and Satt180) associated with yield (Table 16). Satt022 with yield showed dominant effects on the yield as the heterozygote mean was higher than both homozygotes. There were four QTLs which were almost significant ($P \leq 0.07$ to $P \leq 0.11$). These four QTLs were associated with yield (Satt353), maturity (Satt022 and Satt353) and plant height (Satt236).

Comparison of the QTLs in F_2 and $F_{2:3}$ generations of S.100 x CNS population revealed that in F_3 generation, three QTLs were found as in F_2 (Satt180 and Satt353 with yield, and Satt353 with maturity), but, three QTLs were not significant contrary to F_2 generation (Satt022 with yield and maturity, and Satt236 with plant height). Three QTLs, which were not significant in F_2, were detected in F_3. These are Satt050 with yield, and Satt022 and Satt050 with protein which were near to the significance level.

The TN 5-85 x Hutcheson showed low but more consistent SSR marker associations between yield and Satt022, oil content and Satt353 and plant height and Satt353 (Table 15). Satt022 showed a significant dominant effects with yield in F_2 as in CNS x S.100 population. These three QTLs were confirmed in F_3 generation. In addition, another QTL of Satt050 with oil was found in F_3.

Comparison of two population in relation to the QTLs, which were present in both F_2 and F_3, revealed that populations were different for QTLs. This result shows that some QTL associations with agronomic trait were not detected, but other QTLs were formed in soybeans over 60 years. Selection in the long term and involvement of other lines in the pedigrees of modern soybean cultivars might be the reason for the new combination of genes and rearrangements of significant QTLs in the soybean genome. Rasmusson and Phillips (1997) reviewed the research on changes in genetic diversity in the long term and concluded

Table 14. SSR markers associated with soybean agronomic traits and trait means for marker classes in S.100 x CNS population representing ancestral lines

	F_2 plants					$F_{2:3}$ lines			
			Classes					Classes	
Marker	R^2	P	AA	Aa	aa	R^2	P	AA	aa
			Yield (g plant^{-1})					**Yield (g plot^{-1})**	
Satt022	0.05	0.07	66.6	94.4	82.3	0.00	0.97	379.9	381.4
Additive		0.26							
Dominant		0.03							
Satt050	0.02	0.43	76.2	85.8	93.3	0.07	0.06	357.0	439.1
Additive		0.19							
Dominant		0.91							
Satt180	0.05	0.09	99.8	87.4	67.9	0.02	0.34	414.6	361.5
Additive		0.03							
Dominant		0.71							
Satt353	0.04	0.16	70.6	91.1	91.3	0.09	0.03	453.4	348.7
Additive		0.10							
Dominant		0.27							
			Protein (%)					**Protein (%)**	
Satt022	0.01	0.99	38.5	38.5	38.6	0.07	0.08	43.2	42.2
Satt050	0.00	0.83	38.0	38.8	38.6	0.05	0.09	43.0	42.0
			Maturity (day)					**Maturity (day)**	
Satt022	0.03	0.26	160	158	156	0.01	0.56	137	135
Additive		0.11							
Dominant		0.97							
Satt353	0.03	0.25	156	158	159	0.05	0.13	139	135
Additive		0.11							
Dominant		0.59							
			Plant height (cm)					**Plant height (cm)**	
Satt236	0.04	0.12	101.7	99.7	89.3	0.00	0.99	82.9	82.9
Additive		0.07							
Dominant		0.41							

Table 15. SSR markers associated with soybean agronomic traits and trait means for marker classes in TN 5-85 x Hutcheson population representing the fifth cultivar generation.

	F_2 plants					$F_{2:3}$ lines			
			Classes					Classes	
Marker	R^2	P	AA	Aa	aa	R^2	P	AA	aa
			Yield (g plant^{-1})					**Yield (g plot^{-1})**	
Satt022	0.07	0.02	102.2	81.3	109.5	0.15	0.02	557.4	431.8
Additive		0.59							
Dominant		0.01							
			Oil (%)					**Oil (%)**	
Satt050	0.02	0.36	22.3	22.7	22.7	0.06	0.10	21.9	22.3
Additive		0.18							
Dominant		0.50							
Satt353	0.06	0.04	22.4	22.6	23.0	0.11	0.02	21.8	22.3
Additive		0.01							
Dominant		0.40							
			Plant height (cm)					**Plant height (cm)**	
Satt353	0.08	0.01	63.6	62.3	70.1	0.11	0.02	61.1	69.1
Additive		0.07							
Dominant		0.41							

that the genome of crops is more plastic and amenable to selection than previously assumed. Our result support their conclusion. It seems that the Satt022 QTL with yield were transmitted from ancestral lines to modern cultivars in southern soybean breeding lines. This QTL with dominant effect was detected in F_3 in the cross between ancestral lines but it accounted 15 % of the variation for the yield in F_3 in the cross between the modern cultivars.

QTLs are affected by the environment because one side of the association is the expression of the quantitative trait. Even though markers are free of environmental effects, their associations with quantitative traits are subject to environmental changes. Lee *et al.*

(1996) found 11 QTLs for plant height and eight for lodging. Only two QTLs for plant height and one for lodging were detected across locations. In this study, F_2 and F_3 generations were grown in different years and both generations had different linkage relationships. Therefore, the consistency or inconsistency of QTLs over generations is expected. The variance accounted for a given trait by the QTL was increased from F_2 to F_3. This indicates the importance of RIL in the detection of QTL compared to the early generations.

None of the studies published on soybean QTL have used four SSR markers which were utilized in this study. Most of the QTL studies were conducted with RFLP which similar to SSR in term s of inheritance. In both marker types three marker classes in segregating populations can be distinguished (Figure 20). RFLP markers are 12 years older than SSR markers. Therefore, QTL studies were carried out with RFLP markers in the past. In the near future, SSR markers can be utilized in QTL research because of its easiness, low cost, codominance inheritance and other advantages over most of the marker technologies. Thus, we can not make a direct comparison of our results with other studies. However, we summarized the QTLs on the linkage groups where four SSR loci were mapped. Satt022 is on the N linkage group on the USDA/Iowa State University map and Mansur *et al.* (1996) reported the only QTL for leaf length in the C1 linkage group where Satt180 was mapped. The reported QTLs on this linkage group are plant height, maturity and lodging (Lee *et al.* 1996a), protein (Lee *at al.* 1996b and Brummer *et al.* 1997), Seed weight (Mian *et al.* 1996), and water use efficiency (Mian *et al.* 1998). Satt050 and Satt236 are in the A1 linkage group. Lark *et al* (1994), Mansur *et al.* (1996) and Brummer *et al.* (1997) reported QTLs for protein and oil content in this linkage group. In our study, Satt050 showed indications for QTL for protein in S.100 x CNS and for oil in TN 5-85 x Hutcheson. Linkage group H in which Satt353 resides showed QTL with leaf width and leaf area (Mansur *et al.* 1996), and with protein (Brummer *et al.* 1997). However, Satt353 in our study showed association with yield in S.100 x CNS and with plant height in TN 5-85 x Hutcheson populations.

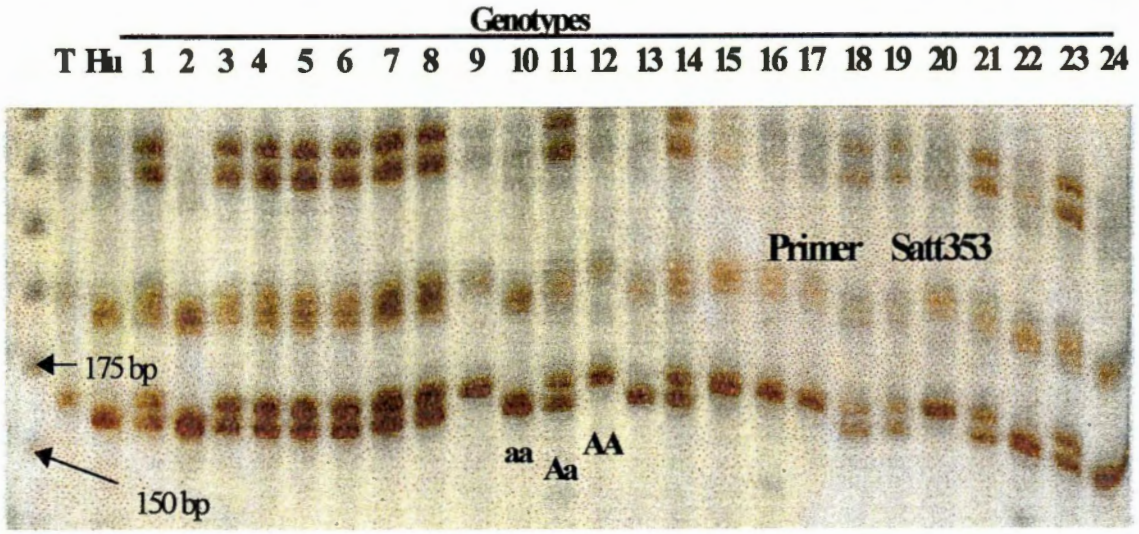


Figure 20. The gel showing segregation for SSR primer Satt353 in TN 5-85 x Hutcheson F_2 plants (genotypes).
T= TN 5-85 and Hu= Hutcheson

CHAPTER 4

Conclusions

Knowledge of the existing variation for agronomic traits is very helpful to set the objectives in plant breeding programs and to decide the selection method for genetic progress. The genetic variability trend over time can also serve to decide breeding strategies. Although, different approaches can be attempted to show genetic variability over time for agronomic trends, narrow sense heritability estimates seem to be more reliable than other biometrical parameter estimation. Recently developed molecular markers may also give accurate results on this matter. Particularly, the associations between molecular markers and quantitative traits (QTL) might better show the genetic variability than biometrical approaches due to less genotype x environment effects and experimental errors.

Although, F_2 data showed high broad sense heritabilities for the yield in southern x southern and northern x southern crosses, the data from F_3 lines derived from F_2 lines did not reflect the variation for yield particularly in southern x southern crosses in this experiment where F_2 plants and $F_{2:3}$ lines were grown in different years. There were a declining trend in genetic variation in F_2, and the remaining variation for yield was sufficient to make genetic progress. However, the same trend was not observed in F_3 due to environmental effects. Genetic variability for protein content and oil content did not exhibit any declining trend. Either trait can be increased within the southern soybean breeding lines. Maturity and plant height had high heritabilities. Heritabilities for maturity in southern x southern populations indicated the decline; however, the remaining variation was still significant. Variability for plant height showed a very small increasing trend. Lee x Ogden population is very atractive to increase genetic variability in southern soybean. Lee and Ogden have four different ancestral lines on the background. Its significant and high heritabilities for all traits makes it very exceptional population for future use. On the other hand, the southern soybean breeders

should practice recurrent selection in northern x southern crosses to increase genetic variability for yield for long term purposes. Although northern x southern crosses showed significant and high heritabilities for yield, their means were lower than southern x southern crosses. Recurrent selection can increase the mean and keep the genetic variability for yield.

Determination of genetic variability over sequential generations of cultivars by QTLs has not been attempted. Two crosses between ancestral lines, S.100 x CNS, and between fifth generation, half sib cultivars, TN 5-85 x Hutcheson were compared for the QTLs. The two fifth generation cultivars are related to the ancestral lines and each other in genetic background. TN 5-85 and Hutcheson have a common parent (Essex). In a cross created by very close parents, QTLs with yield, oil and plant height were detected. Another QTL was on the border for significance in the same population. This result points out the usefulness of SSR markers in molecular studies.

QTLs present in the two ancestral lines were not detected in the fifth generation cultivars, and the QTLs present in the modern cultivars were not found in the two ancestral lines. The dynamics of plant genome should form new QTLs with the effect of selection and new parent combinations. The similarity in the number of significant QTLs between ancestral lines and modern cultivars imply that the complexity of plant genome has tools to create new variability while losing some variability. In the southern gene pool, genetic variability has been reduced to a degree, but on the other hand, QTLs in this study shows that there is remaining genetic variability which cannot be detected by biometrical approaches.

The comparison of QTL results and heritability results in this study indicates that QTL analysis can be more sensitive to estimate genetic variability. By the heritability estimation, the genetic variation for yield in S.100 x CNS population was very low and non-significant and genetic variation for yield in TN 5-85 x Hutcheson cross was zero. The same yield data was analyzed for QTLs and genetic variation was detected for different traits in both populations. However, some inconsistencies between generations for the QTLs in this study point out the problems of confirming in succeeding generations, QTLs found in one

populations and in one environment. The increase of knowledge on QTLs, advancements in molecular marker technologies, and increasing the number of QTLs in soybean can open new doors in plant breeding.

LITERATURE CITED

Allen, F. L. and H. L. Bhardwaj. 1987. Genetic relationships and selected pedigree diagrams of North American soybean cultivars. The Univ. of Tennessee, Agricultural Experiment Station. Bulletin 652. Knoxville, TN

Bailey, M. A., M. A. R. Mian, T. E. Carter, Jr, D. A. Ashley, and H. R. Boerma. 1997. Pod dehiscence of quantitative trait loci. Heredity 88(2):152-154

Bassam, B. J., G. Caetano-Anollés, and P. M. Gresshoff. 1991. Fast and sensitive silver staining of DNA in polyacrylamide gels. Anal. Biochem. 196:80-83

Bolstein, D., R. L. White, M. Skolnick, and R. W. Davis. 1980. Construction of a genetic map in man using restriction fragment length polymorphism. Am. J. Hum. Genet. 32:314-331

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

Burton, J. W. 1987. Quantitative genetics: Results relevant to soybean breeding. *In* J. R. Wilcox (Ed.) Soybeans: Improvement, production and uses. Second edition. ASA, CSSA, SSSA, Madison WI. p. 211-247

Carlson, J. E., L. K. Tulsieram, J. C. Glaubitz, V. W. K. Luk, C. Kauffeldt, and R. Rutledge. 1991. Segregation of random amplified DNA markers in F_1 progeny of conifers. Theor. Appl. Genet. 83:194-200

Delaney, X., D. M. Rodgers, and R. G. Palmer. 1983. Relative genetic contributions among ancestral lines to North American soybean cultivars. Crop Sci. 23:944-949

Dellaporta, S. L., J. Wood, and J. B. Hicks. 1983. A plant DNA minipreparation. Version II. Plant Mol Biol. Rep. 1:19-21

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

Dudley J. W., and R. H. Moll. 1969. Interpretation and use of estimates of heritability and genetic variances in plant breeding. Crop Sci. 9:257-262

Falconer, D. S. and T. F. C. Mackay. 1995. Introduction to quantitative genetics. 4th ed. Longman Group, Ltd., Essex, England

Gizlice, Z., T. E. Carter, Jr, and J. W. Burton. 1993a. Genetic diversity in North American soybean: I - Multivariate analysis of founding stock and relation to coefficient of parentage. Crop Sci. 33:614-620

Gizlice, Z., T. E. Carter, Jr, and J. W. Burton. 1993b. Genetic diversity in North American soybean: II - Prediction of heterosis in F_2 populations of southern founding stock using genetic similarity measures. Crop Sci. 33:620-626

Griffin, J. D., and R. G. Palmer, 1987. Inheritance and linkage studies with five isozyme loci in soybean. Crop Sci. 27:885-892

Jones, C. J., K. J. Edwards, S. Castaglione, M. O. Winfield, F. Sala, C. van de Wiel, G. Bredemeijer, B. Vosman, M. Matthes, A. Daly, R. Brettschneider, P. Bettini, M. Buiatti, E. Maestri, A. Malcevschi, N. Marmiroli, R. Aert, G. Volckaert, J. Rueda, R. Linacero, A. Vazquez, and A. Karp. 1997. Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. Molecular Breeding 5:33-42

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

Lark, K. G., J. Orf, and 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. 88:486-489

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

Lee, S. H., M. A. Bailey, M. A.R. Mian, T. E. Carter, Jr, E. R. Shipe, D. A. Ashley, W. A. Parrott, R. S. Hussey, and H. R. Boerma. 1996b. RFLP loci associated with soybean seed protein and oil content across populations and locations. Theor. Appl. Genet. 93:649-657

Lewers, K. S., E. H. Crane, C. R. Bronson, J. M. Schupp, P. Keim, and R. C. Shoemaker. 1999. Detection of linked QTL for soybean brown stem rot resistance in 'BSR 101' as expressed in a growth chamber environment. Molecular Breeding 5:33-42

Liu, B. H. 1997. Computational tools for study of complex traits. *In* A. H. Paterson (Ed) Molecular dissection of complex traits. CRC Press, p. 43-81

Mansur, L. M., J. Orf, and 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

Mansur, L. M., J. H. Orf, K. Chase, T. Jarvik, P. B. Cregan, and K. G. Lark. 1996. Genetic mapping of agronomic traits using recombinant inbred lines of soybean. Crop Sci. 36:1327-1336

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

Mian, M. A. R., D. A. Ashley, and H. R. Boerma, 1998. An additional QTL for water use efficiency in soybean. Crop Sci. 38:390-393

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

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

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

Skorupska, H. T., R. C. Shoemaker, A. Warner, E. R. Shipe, and W. C. Bridges. 1993. Restriction fragment length polymorphism in soybean of the southern USA. Crop Sci. 33:1169-1176

St. Martin, S. K. 1982. Effective population size for the soybean improvement program in maturity groups 00 to IV. Crop Sci. 22:151-152

Tautz, D. 1989. Hypervariability of simple sequences as a general source of polymorphic DNA markers. Nucl. Acids Res. 17:6463-6471

Wricke, G. and W. E. Weber. 1986. Quantitative genetics and selection in plant breeding. Walter de Gruyter, Berlin New York

APPENDICES

Appendix 6. Narrow sense heritabilities (h^2) and standard errors (Se) for heritabilities by parent-offspring regression for maturity and plant height in different F_2 and $F_{2:3}$ soybean populations representing different eras in soybean breeding.

Population	h^2	Se	Population	h^2	Se
Southern x Southern			**Northern x Southern**		
			Maturity		
S.100 x CNS	0.85**	0.10	Mandarin x CNS	1.46**	0.11
Lee x Ogden	0.79**	0.15	Harosoy x Lee	1.47**	0.14
Essex x Unknown	0.97**	0.28	Williams x Essex	1.07**	0.14
TN 5-85 x Hutcheson	0.39**	0.13	BSR 101 x TN 5-85	0.94**	0.11
			Plant height		
S.100 x CNS	0.05	0.10	Mandarin x CNS	0.70**	0.06
Lee x Ogden	1.02**	0.10	Harosoy x Lee	1.06**	0.07
Essex x Unknown	0.79**	0.12	Williams x Essex	1.01**	0.06
TN 5-85 x Hutcheson	1.07**	0.13	BSR 101 x TN 5-85	1.09**	0.08

** significant ($P \leq 0.01$)

PART IV

SUMMARY
AND
FUTURE RESEARCH

Summary

Evaluation of genetic variability in southern soybean is needed to plan future breeding strategies and understand how genetic variability changes in soybean breeding lines in the long term. This study involved performance and evaluation of selected ancestral lines and unique cross combinations between ancestral lines, and first, third and fifth generation cultivars.The selected ancestral lines and cultivars were chosen because:

(1) they represented different eras of production from 1940's to 1990's,

(2) the selected ancestral lines are major contributing lines to the modern gene pool of North American soybean cultivars,

(3) the cultivars represent different generations of related cultivars and they were widely grown during their era of production, and

(4) most of them have been utilized extensively as parents in both public and private breeding programs.

Eight southern ancestral lines and cultivars from first, third and fifth generation cultivars (CNS, S.100, Ogden, Lee, Hill, Essex, TN 5-85 and Hutcheson) were tested in four locations in Tennessee. Experiments were conducted in Knoxville and Ames Plantation for 3 years and in Milan and Springfield for 2 years in a randomized complete block design with three replications. The estimated genetic progress for soybean yield in the mid-southern states has increased 14 kg $ha^{-1}yr^{-1}$. This increase is comparable to yield increases in the northern states and MG V-VII in the southern states. The yield increase was linear which implies that there is still remaining variation in southern soybean gene pool to make progress for yield. Oil content was increased until 1980's, but started to decrease after that. Because negative relationship between oil and protein content, protein content was decreased first and then increased in this period. Lodging resistance was also increased in soybeans. Plant height followed a similar trend to oil content. Maturity was decreased linearly.

The crosses were made to represent ancestral lines, and first, third and fifth generation

cultivars. Twenty F_1 seeds from each cross were grown in Costa Rica. F_2 rows from each F_1 plants and parent rows were planted in a incomplete block design in Knoxville. Yield, protein and oil content, plant height and maturity were recorded in 250 randomly selected F_2 plants from each cross and 50 plants from each parent. Narrow sense heritability for yield, protein and oil content showed a declining trend. Molecular marker similarities were estimated for the total of 12 parents. DNA markers (SSR and DAF) showed also a declining trend. However, molecular marker similarities were not correlated with heritabilities for agronomic traits with the exception of correlation between SSR-GS and heritability for yield. Due to small sample size (n=8), the correlation between SSR-GS and yield heritability needs to be confirmed. Coefficient of parentage and quantitative genetic similarities were also estimated. Both estimates were not correlated with yield. It was shown that two assumptions of coefficient of parentage can not be held. The first assumption is there is no relationship between unrelated genotypes in pedigree. DAF-GS showed that the similarity between unrelated genotypes around 50 %. The second assumption is equal contributions of each parent to their progenies. SSR-GS and DAF-GS indicated in this study that this assumption is no longer valid. Correlations among SSR-GS, DAF-GS, CP and QT-GS were significant. The CP estimation should be improved with the help of DNA markers by correcting two assumptions.

Randomly selected F_2 plants in each of the eight crosses were advanced to F_3 generation and agronomic traits were recorded as in F_2 generations. $F_{2:3}$ lines from each population and parents were planted in rows in an incomplete block design in Knoxville. Parent-offspring regression was applied to estimate narrow sense heritabilities. There was no significant heritability for yield except in Lee x Ogden population in southern x southern populations, but they were significant in northern x southern populations. Heritabilities for protein and oil content were similar or equal in southern x southern and northern x southern populations. They were not significant in southern ancestral populations. The distribution of yield in F_2 and F_3 were similar showing that the mean of southern x southern crosses was

higher than that of northern x southern crosses. Lee x Ogden gave the highest mean among all populations. This particular cross showed also exceptionally high genetic variations for yield, protein and oil content, maturity and plant height. It can be used as a source material in southern breeding programs. It is also very suitable for studies on classical and molecular genetics, because it has significant and high heritabilities for all quantitative traits in this study, four different ancestral lines on the genetic background and highest yield mean among all populations in F_2 and F_3 generations.

S.100 x CNS and TN 5-85 x Hutcheson populations representing ancestral lines and fifth generation cultivars, respectively, were also evaluated for SSR marker associations with agronomic traits. In S.100 x CNS, SSR markers Satt022, Satt180 and Satt353 were associated with yield in F_2 but Satt022 was not confirmed in $F_{2:3}$ lines. There were QTLs for other traits which were either significant or near to significant level either in F_2 or F_3. In TN 5-85 x Hutcheson population, Satt022 was associated with yield and Satt353 was associated with oil content and plant height. The change in the QTL relationships from ancestral lines to the fifth generation cultivars implies that some genetic variability may have been lost or the association of the marker with QTL was disrupted due to crossover(s), but some other variability was gained over time in southern breeding lines. The variability for yield, which was not detected by heritabilities was found by SSR marker QTLs. This indicates that QTLs can be more sensitive way to demonstrate genetic variability.

Future Research

This study also gave indications for the reseach needs highlighted below. Proposed research areas may give useful results to advance soybean breeding in southern states as well as in other areas. The results may also serve plant breeding programs in general.

1- Two assumptions of the coefficient of parentage can be investigated by markers in which arbitrary primers are employed. Minimum similarity between breeding lines can be used as a coefficient between individuals unrelated in pedigree. It seems that the ancestral lines (around 40-50) can serve for this purpose.

2- Use of markers in parent choice for crosses can be studied by the markers with QTL effect. SSR markers can be better for this purpose because of low cost, high polymorphism and easy screening. However, the number of SSR markers with QTL is not sufficient. In the next 2-3 years time, it is expected that there would be sufficient number of SSR markers with QTL effect.

3-A study is needed on cluster analysis to show which cluster method is the best for cultivar classification and what kind of data (raw data or secondary data like genetic similarities) should be used as an input.

4- Studies on QTL changes over cultivar generations should be continued with more populations representing ancestral lines, first, second, third, fourth, fifth and six generation cultivars. It may give better idea about genetic variability changes. Also it may supply information which part of chromosomes carry the desirable genes.

5- Recombinant inbred lines of Lee x Ogden population should be developed. In F_6 generation, sublines from each recombinant inbred lines can be established based on maturity, plant height and pubescence color. These inbred lines can be used for a collaborative research on QTL mapping. Parallel to QTL mapping, quantitative genetic studies can be conducted with these lines and the lines in different generation to be created by crossing Lee and Ogden. These lines also can be utilized in soybean breeding programs for cultivar improvement.

VITA

I was born on March 10,1961 in Amasya, Turkey. I finished elementary and middle school there. I attended a specific highschool for agriculture in Tokat, Turkey and finished in 1978. I got my first degree from 19 Mayis Universitesi in 1983 in Samsun, Turkey.. At the same university, I graduated from M.Sc. program in 1986. I have worked as a legume breeder for 4 years in The Turkish Ministry of Agriculture. The Ministry of Agriculture sent me to Wales, UK. to study M.Sc.in plant breeding. I graduated from master program in 1992 at the University of Wales, Aberystwyth, UK.

After the master study in Wales, I started to work as department head for the plant breeding department in the Black Sea Agricultural Research Institute, Samsun, Turkey. I was offered PhD studies in the U.S. by the Turkish Ministry of Agriculture and I first went to the University of Florida. Then I was transferred to the University of Tennessee in 1995. I worked with Dr Allen for my PhD, and deserved PhD in May 19, 1999. I have been married with Meral for 11 years. I have a 10 years daughter, Damlanur, and one year old son, Alptekin.