ABSTRACT

RECKER, JILL RENEE. Analyzing Variances and Correlations of Quantitative Traits in Two Long Term Randomly Mated Soybean Populations. (Under the direction of Joseph W. Burton).

Long term random mating and natural selection in two soybean [Glycine max (L.) Merr.] populations revealed pleiotropic causes of well known genetic correlations. Genotypic variances, heritabilites and correlations as well as phenotypic variances and correlations were evaluated to determine the effects of 26 generations of random mating. Yield plots were grown in three locations over one year with two replications per location. Data was collected on flower color, flowering date, maturity date, pubescence color, height, lodging, yield, seed weight, percent protein, and percent oil.

RSII had a yield mean of 2070 kg/ha with a mean protein of 41.61% and a mean oil of 19.33%. Heritability estimates, on an entry mean basis, were high (>80%) for flowering date, maturity date, lodging, and height. Heritability for yield was estimated at 33.1%, protein at 68%, and oil at 56%. The genetic correlation between percent protein and yield was negative but not significant. Percent protein and percent oil were significantly negatively correlated. Percent oil and yield also had a non-significant positive correlation. Theoretical responses to selection for yield, oil, and protein were predicted to increase by 161 kg/ha, 0.57%, and 1.0% respectively. Selection for yield is predicted to slightly increase oil 0.03% and decrease protein by 0.12%.

RSIII had a yield mean of 2205 kg/ha with mean protein of 41.61% and mean oil of 19.04%. Heritability estimates, on an entry mean basis, were high (>80%) for flowering date, maturity date, height, seed weight, protein, and oil. The heritability estimate for yield was 55%. The genetic correlation between yield and percent protein was non-significant and positive. Percent protein and percent oil had a significant negative genetic correlation. Yield and percent oil had a non-significant positive correlation. Theoretical responses to selection for yield, oil, and protein were predicted to increase by 272 kg/ha, 1.47%, and 1.96% respectively. Selection for yield is predicted to slightly increase both protein and oil by 0.11% and 0.29% respectively.

Considering the long term random mating in the populations, genetic linkage disequilibrium has decreased significantly and any genetic correlation can be assumed to be caused by pleiotropy. Given the significant genetic variances in these two populations and high heritabilities, random mating followed by inbreeding has been an effective way to produce high yielding inbred lines for use in other breeding experiments.

Analyzing Variances and Correlations of Quantitative Traits in Two Long Term Randomly Mated Soybean Populations

by
Jill Renee Recker

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Crop Science

Raleigh, North Carolina

2010

APPROVED BY:

_____	_____
Andrea J. Cardinal	David A. Dickey

_____
Joseph W. Burton
Chair of Advisory Committee

DEDICATION

I would like to dedicate this to my parents, who instilled within me a passion for agriculture, gave me the drive to succeed, and allowed for my curiosity to flourish. Also to my family and friends who have kept me grounded throughout the years and who acted as outlets during life's stressful times.

BIOGRAPHY

I grew up on a family farm in Genoa, OH where I was highly involved in every facet of that lifestyle. During that time I was active in both 4-H and FFA which allowed for me to investigate agriculture beyond the farm. Once at Purdue University, I was involved in the Agronomy Club as well as within the Agronomy Department researching under Dr. Herb Ohm. I interned at Monsanto, HortResearch, and AgRelaint during my university career as well. Currently I am a student at North Carolina State University where I will further pursue my Plant Breeding career by working towards my PhD in the Crop Science Department.

ACKNOWLEDGEMENTS

I would like to thank Dr. Joe Burton for all his advice and knowledge he shared during my experience working with him. Also I want to thank the staff at the soybean unit for their help during planting, harvesting, and every other step along the way as well as their advice and excellent field experience. I would also like to thank my committee for their helpful advice and reviews of this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ANALYZING VARIANCES AND CORRELATIONS OF QUANTITATIVE TRAITS IN TWO LONG TERM RANDOMLY MATED SOYBEAN POPULATIONS

Introduction

Soybean [*Glycine* max (L.) Merr] is one of the most important crops in the world's oilseed and protein industries, accounting for 90% of the oilseed market and providing two-thirds of the world's protein (Ash, Dohlman, & Wittenberg, 2008). Soybean meal also accounts for 50-75% of its total value. From 1985 to 2004, US soybean production increased to 75 million Mg and in 2004 accounted for 45% of the world's soybean production. Many US states had record yields from 2004 to 2006. Soybean production is expected to increase worldwide over the next decade although the US acreage may decrease during that time. In 2008, the US average yield was 2661 kg/ha with yields rising an estimated 30.2 kg/ha per year (based on data from 1960 to 2007) (Ash & Dohlman, 2008).

Soybean cultivars with high levels of protein, oil, and yield are needed to sustain current and future demands. Wilson (2004) noted that to optimize the value of soybeans, protein levels should be between 44 to 45% and oil levels no less than 18%, on a dry matter basis. This is not common in current cultivars. In 2008, the average protein and oil contents across the Midwest were 40.9% and 20.6% respectively, according to the United Soybean Board. Within the USDA germplasm collection the average protein and oil content was 42.1% and 19.5% respectively (Wilcox, 2004).

Yield, Protein, and Oil Improvement

Yield has seen significant improvements over the last century. Boerma's (1979) study of soybeans grown in the southern US (maturity group VI, VII, and VIII) indicated that an average 0.7% yield increase per year was observed from 1942 to 1973. Wilcox (2001) reviewed the 2 year regional testing average for the 3 highest yielding cultivars in each maturity group (00 to IV) from 1951 to 1999. He found significant yield gains across all evaluated maturity groups, having at least a one percent gain in yield per year. Yield gains may have been limited due to the fact that many current soybean cultivars have closely related pedigrees. Gizlice, Carter, and Burton (1994) estimated the total number of ancestors that gave current cultivars their genetic base. Of the southern varieties developed between 1947 and 1988, 17 ancestors made up 94% of their genetic base. They also estimated that of the northern varieties developed from 1947 to 1988, only 10 ancestors constituted 80% of their genetic base. Usten, Allen, and English (2001) determined that even though cultivars of the eighties came from similar backgrounds, a yield plateau had not been reached, in fact they estimated a 20% gain in yield from ancestral lines to cultivars of the eighties.

For many selection experiments, the main objective is to increase yield, although selection for increasing protein and oil are also very important. Progress may be limited though with certain selection methods for any of these traits individually or when selecting for all three traits together depending on the breeding method used. Brim and Burton (1979) claimed that

of the current breeding methods, pedigree and backcrossing are generally ineffective for selection on increased protein content, perhaps from a decrease in genetic variability for protein. A study by Helms and Orf (1998) detailed an experiment of $F_{4:5}$ lines for protein selection and evaluated the indirect selection effects on yield. They compared selection for high protein based on individual line phenotypes, independent of population structure, versus selection based on population means, selection on the whole population, and the correlated response in yield. They expected that selection for high protein among populations will maintain greater genetic variance and have a less detrimental effect of yield. However, both selection experiments observed an increase in protein while yield and oil decreased. Their study was among many that had either yield or oil decreases while selecting for increased protein or vice versa (Brim & Burton, 1979; Usten, et. al., 2001). Given the difficulties in developing a soybean variety with high values of yield, protein, and oil, the development of the cultivar 'Essex' (Smith & Camper, 1973) was thought to be a milestone for breeders and growers in the southeastern states. This is because Essex has both high percent protein and high percent oil with relatively good yield, but was one of a very few cultivars at the time it was released. Other cultivars have been released with significant increased protein, e.g. Prolina, Vinton, and Protana (Burton, Carter, & Wilson, 1999; Bahrenfus & Fehr, 1980; Probst, Laviolette, Athow, & Wilcox, 1971), but all were lower yielding than their contemporary cultivars.

Response to selection for yield, protein, and oil will greatly depend on their heritabilities and genetic correlations with each other. Entry mean heritability estimates from many selection experiments were summarized by Burton (1987). Yield heritability estimates varied from as low as 3% to as high as 58% while protein and oil generally had less variability in their estimates. Protein ranged from 57% to 90% and oil ranged from 51% to 89%. Genetic correlations are useful for breeders to predict indirect selection responses. Burton (1987) also summarized genetic correlations from eight populations. Most populations had positive genetic correlations between yield with plant height, lodging, and days to maturity. The genetic correlation between yield and percent oil was found to be low and generally positive while yield and percent protein had a negative genetic correlation. It is also well known that the correlation between percent protein and percent oil are negatively associated (Brim & Burton, 1979; Usten, et. al., 2001). Thus, in those populations, selection for yield and protein as well as protein and oil would have been fruitless. If these types of genetic correlations and heritabilities are found across most soybean populations, developing a soybean cultivar with high yield, high protein and high oil will be very difficult. In theory, random mating under natural selection will allow for the accumulation of favorable gene combinations and possibly the obtainment of high yield, protein, and oil cultivars if favored under natural selection.

Random Mating's role in Gene Linkages and Pleiotropy

Falconer and Mackey (1996) evaluate genetic correlations for three main reasons: pleiotropic action of genes, changes brought about from selection, and natural selection in terms of fitness. Genetic correlations are caused from either genes in linkage phase disequilibrium or pleiotropy. Linkage phase disequilibrium (LD) is the non random association of alleles at different loci and can be increased from population subdivision, population mixture, mutation, physical linkages, and selection on epistaic interactions. LD can be decreased through genetic recombination and independent assortment of loci. Pleiotropy is caused from the action of a single gene having multiple phenotypic effects in more than one trait. When the phenotypes of two traits are controlled by genes that are in linkage phase disequilibrium (LD), many cycles of genetic recombination will reduce the LD and new gene combinations can exist. If favorable LD is present for certain genetic correlations, the breeder will want to keep those while trying to reduce unfavorable LD correlations. However, if the reason two traits are correlated is due to pleiotropy, random mating will not reduce the genetic correlation between the two traits and they will always be correlated. The distinction and knowledge of what causes genetic correlations is vital since it is the determining factor of whether or not any breeding strategy will be effective in obtaining desirable phenotypes of two traits when selection is being performed on both traits.

Theoretically, long term random mating should allow breeders to distinguish between LD and pleiotropy. Pleiotropy is caused from either direct or indirect factors. Direct implies that the correlated phenotype is caused by the same gene having more than one function. Indirect implies that the phenotype is connected through a physiological process. Pleiotropy can be detected by allele replacement in the particular gene causing the two phenotypes. Once the linkage disequilibrium is reduced to near zero and the population reaches linkage equilibrium, pleiotropic effects can be determined (Hartl & Clark, 2007). This can be a slow process however, dependent on recombination rates. Once the distinction between LD and pleiotropy has been determined, it will allow breeders to develop effective breeding programs to select and obtain favorable trait combinations.

Under an ideal population meeting the assumptions of: distinct generations, the number of individuals in each generation is constant, no migration, no selection, no mutation, and random mating; allele frequencies will not change over time from the initial gene frequencies (Falconer & Mackay, 1996). If any of these assumptions are not met, gene frequencies will change from the initial parental frequencies. In biparental populations, gene frequencies are contributed by only two individuals, making calculations easy and allele frequencies large enough to not be lost quickly through random genetic drift in an ideal population. This makes determining if the population is ‘ideal’ easier as calculations are easier and expectations are likely met when random mating occurs. If a population is developed from multiple parents,

gene frequencies may be small enough to be quickly lost due to random genetic drift. With multiple parent experiments, the development will be more difficult as many more crosses will be needed with a self-pollinated species such as soybeans. Also the calculations and expectations can be more difficult be compute.

Soybean Breeding

Current soybean breeding methods involve mating breeding lines, superior cultivars, and/or exotic germplasm, in 2 or 3 and sometimes 4 way crosses to develop superior progeny. Selection can be made during the inbreeding process or once homozygous lines have been obtained. Inbreeding is rapid, usually starting in the F_2 generation by self pollination. This decreases the amount of genetic recombination possible. During inbreeding, lines are selected and superior genotypes are identified through testing and then released as cultivars. Breeding procedures for soybeans include pedigree selection, bulk selection, single seed descent, backcrossing to a recurrent parent, or recurrent selection with or without the use of genetic male sterility (Orf, Diers, & Boerma, 2004).

Since soybeans are naturally self-fertilized, the principles of recurrent selection are hard to maintain because an intermating step is required in each cycle. Genetic male sterility has therefore proven to be the fastest and least laborious method for random mating in self pollinated species. Genetic male sterility takes advantage of insect pollinators, mainly bees,

to make all the crosses within a population since all seed produced is from a non-self pollen source and therefore hybrid. Orf et. al (2004) reviewed studies since 1984 that have utilized recurrent selection. The most selected trait of interest in recurrent selection was yield. Many other traits have also effectively improved utilizing recurrent selection with genetic male sterility such as oil, protein, seed size, and maturity.

With genetic male sterility it is possible to obtain random mating populations and ease in outcrossing for a normally self-pollinated species. Several studies have shown that random mating increases genetic variances. Kenworthy (1979), evaluated different crossing schemes to identify the benefits of each crossing method when using exotic germplasm. He determined that intermating proved to be the most beneficial method for increasing genetic variance. Silvela and Diez-Barra (1985) agree with Kenworthy that recurrent selection and intermating is beneficial and useful. Their study evaluated artificial selection in a self-pollinated species through mathematical models to compare the influence of recurrent selection on single and multiple locus models versus selfing during selection. They concluded that recurrent selection was advantageous for dominant and over-dominant alleles as well as when negative linkage disequilibrium is present in the population (Silvela & Diez-Barra, 1985). Negative linkage disequilibrium compared to those in Hardy Weinberg Equilibrium have an excess of intermediate genotypes and too few extreme genotypes. Silvela, De La Pena, and Gomez-Ruano (1999) re-evaluated the assumption that random

mating surpasses inbreeding for autogamous species in a study of the beetle *Tribolium castaneum* Herbst. They concluded from that study that random mating was the best method to form extreme genotypes (both good and bad). They also concluded that any benefit in performance was due to additive and epistatic gene actions. Finally, their study revealed that epistatic interactions were far more important than dominance in autogamous species since natural selection is acting on homozygotes and therefore those interactions will be fixed in the homozygote state. Stam (1976) also evaluated the effects of random mating in a computer simulation. He concluded that random mating is preferred for at least one generation, perhaps in the F_4 generation of selfing in a biparental cross. Since recurrent selection includes random mating, it is thought to break both coupling and repulsion phase linkages while at the same time creating new favorable gene combinations.

In a study published by Burton, Koinange, and Brim (1990), the same two populations in this study, RSII and RSIII, were evaluated for their response to selection of S_1 families for yield after 4 cycles of recurrent selection. RSII had an increase in yield from 1840 to 2222 kg/ha. Burton et. al. (1990) found an increase in plant height and a decrease in lodging through the correlated response to selection on yield. There was significant yield gain in RSIII over three cycles, but lower than RSII, and no gain over four cycles. In RSIII, there were also no significant changes in other traits measured. The success to selection in RSII was attributed to the differences in genetic variability and heritability. They found genetic variability in

RSII to be twice that in RSIII and heritability estimates to be three to four times greater in RSII as well.

Tinius, Burton, and Carter (1991) conducted a selection study using three sub-populations of RSIII after nine cycles of random mating, which was followed by three and four cycles of recurrent mass selection for increased seed size in male sterile plants. They found an increase in both male sterile and male fertile seed size. An increase in seed yield was also observed after the four cycles of selection. Tinius, Burton, and Carter (1993) evaluated the effect of seed size selection on seed composition traits in the same populations, and found a decrease in percent protein and an increase in percent oil. Negative phenotypic correlations were reported between percent protein and seed size as well as percent protein and percent oil.

The two populations, RSII and RSIII, studied by Burton et. al. (1990) were randomly mated for additional 18 generations to produce the two soybean populations of this study. They were developed at NCSU and utilized genetic male sterility to produce out-crossing soybean populations. These populations were randomly intermated without selection for a total of 26 generations by harvesting seed only from the male sterile plants and replanting that seed the following year. Lines were developed from 6 generations of inbreeding, through single seed decent, and were evaluated in yield trials in the $F_{5:7}$ generation. The only selection performed on these lines during the inbreeding process was to discard genetic male sterile plants.

Currently the RSII population (RSII) contains 119 lines and the RSIII population (RSIII) has 103 lines. With 26 generations of random mating without selection, these populations are expected to be in linkage equilibrium at most pairs of loci and if significant genetic correlations are observed they could be assumed to be caused by pleiotropic effects or being favored under natural selection.

To date, little is known on how random mating in soybeans will affect population parameters. Also it is unknown how much pleiotropy affects breeding methods and population estimates in comparison to linkage disequilibrium. The objective of this study was to determine how populations parameters, means, genotypic and phenotypic variances, and covariances, were affected by 26 cycles of random mating. Genetic variances, genotype by environment interactions, genotypic and phenotypic correlations and covariances, and heritabilites were determined. These were compared with those in the literature to determine what impact, if any, random mating has had on correlations between traits, particularly, those between seed composition and yield. A second objective was to compare productivity of these materials with current elite cultivars to determine if random mating alone had provided useful germplasm for soybean breeders. Finally, the third objective was to determine what, if any, implications these populations reveal for current soybean breeding methods.

Materials and Methods

Population Development

The two populations in this study, designated RSII and RSIII, were developed by the North Carolina Agricultural Research Service and UDSA-ARS. RSII was developed by continuous random mating, with no selection, of the registered population N79-1400 (Reg. No. GP 40) for an additional 18 generations for a total of 26 generations. N79-1400 was developed from crossing the male sterile maintainer line, N69-2774 (Brim and Young, 1972) to 10 of the highest yielding lines from the population, CY2, which was in its second cycle of a recurrent selection. The CY2 experiment began with a cross between D49-2491 and nine plant introductions, which were unadapted to southern USA regions and had high levels of percent seed protein. These F_1's were then backcrossed to D49-2491. The nine plant introductions were: PI 69035, PI 165673, PI 54618-1, PI 31649, PI31676, PI 97100, PI 153681, and PI 171445 (Burton & Brim, 1981; Appendix B1).

RSIII was developed by continuous random mating, with no selection, of the registered population N79-1500 (Reg. No. GP 41) for an additional 18 generations for a total of 26 generations. N79-1500 was developed from a cross between six cultivars/experimental lines and the male sterile maintainer line, N69-2774. The six male lines were well adapted to the southern USA regions and most showed resistance to a number of major plant diseases (bacterial pustule, target spot, frogeye leaf spot, root-knot and soybean cyst nematode), as

well as one cultivar with high percent oil and two cultivars with high percent protein. The six male parents were: ‘Ransom’, ‘Govan’, D65-6765, D68-216, F66-698, and N67-4063 (Burton & Brim, 1981; Appendix B2).

Once the populations reached 26 total cycles of random mating, with no selection, each population was inbred from single seed descent to the F_4 generation with selection against male sterility and single plants were harvested. $F_{4:5}$ plants were grown in single plant rows in Clayton, NC in the fall of 2008 and a single plant was harvested from each row with selection against those rows exhibiting male sterility as well as those rows with outlying maturities, which were believed to be outcrosses. In the winter of 2008/2009 the $F_{5:6}$ seed was grown in Puerto Rico to increase seed for yield testing, and seed was harvested by row in bulk. The seed harvested from Puerto Rico yielded 119 lines of RSII and 103 lines of RSIII.

Given the population development, it is expected that half of the genetic makeup in both populations would consist of the male sterile maintainer cultivar in the initial generation. Over time, within each population, the genetic makeup of the male sterile maintainer genotype would change with the population, while still contributing half of its genetic makeup to its progeny. Direct comparisons between the two populations cannot be made though since they were grown and tested in two different experiments.

Experimental Procedure

In the summer of 2009, the $F_{5:7}$ lines of each population were planted at three North Carolina locations: Mount Olive, Plymouth, and Clinton with two replications at each location. The planting dates for the three locations were June 1, 2009, May 27, 2009, and May 20, 2009 respectively. 'NC-Raleigh' was used as a check cultivar in the RSII experiment and both 'NC-Raleigh' and 'NC-Roy' were used as check cultivars in the RSIII experiment. The genotypes were planted in a Randomized Complete Block. Parents were not included because seed was not available at the time of trials. Three row yield plots were grown at Clinton and Plymouth that were 6.7m long with 0.97m spacing between each row. Plots were end trimmed to 4.9m in the R7 stage of development. Four row yield plots were grown at Mount Olive that were 4.9m long with no end trimming and 0.97m spacing between each row. The soil type at the three locations were 'Dragston Loamy Sand' (Coarse-loamy, mixed, semiactive, thermic Aeric Endoaquults) at Mount Olive, 'Portsmouth' (fine sandy loam or sandy skeletal, mixed, thermic, Typic Umbraquult) at Plymouth, and 'Orangeburg Loamy Sand' (Fine-loamy, kaolinitic, thermic Typic Kandiudults) at Clinton.

During the 2009 growing season, phenotypic data was collected for flower color, flowering date, maturity date, pubescence color, height, lodging, and yield. Flowering date and maturity date were taken from July 1 and October 1 respectively and considered flowering when 95%

of the row had flowered and mature when 95% of the pods were mature. The center row/s of each yield plot was harvested on January 11, 2010 at Mount Olive, December 22, 2009 at Plymouth, and November 28, 2009 at Clinton. 100 seed weights were taken from each harvested plot after ambient air drying. A random subsample of 25g was sent from each yield plot to the Functional Food Research Unit in Peoria, IL to obtain percent protein and percent oil, on a dry matter basis, from their near infrared Infratec® 1255 Whole Grain Analyzer.

Statistical Analysis

Genotypic variances, covariances, correlations, and heritabilities were calculated as well as phenotypic covariances and correlations for eight traits: yield, seed weight, plant height, maturity date, flowering date, lodging, percent protein, and percent oil.

An analysis of variance was performed using the SAS system's procedure for fitting general linear models, PROC GLM (SAS Institute 2003). All variables in the model were treated as random effects. Each model was determined to be adequate by looking at the overall F-test, (for which a specific F-test was given: test h=geno e=loc*geno), R-squared value, and Coefficient of Variation. Expected mean squares were determined for each source of variation and are shown in the Table 1. The coefficient for each variance component are listed as symbols for

generalizing. Tests for normality in each population were made with PROC UNIVARIATE in the SAS system (SAS Institute 2003) and found normal.

Genotypic variances were estimated by using the SAS system's procedure for estimating the random effects variance components (PROC VARCOMP) (SAS Institute 2003). Standard Errors were calculated for each genotypic variance using the equation suggested by Anderson and Bancroft (1952).

Genetic covariances for all combinations of traits were calculated by using the output from the MANOVA statement in SAS under PROC GLM (SAS Institute 2003):

```
manova h=geno e=loc*geno/printh printe;
```

The output was used to determine the Mean Cross Product and then to obtain covariances:

$$\frac{\left[\left(\frac{H = Type\ III\ CP\ for\ each\ pair}{degrees\ freedom\ for\ geno(block)}\right) - \left(\frac{E = Type\ III\ CP\ for\ each\ pair}{degrees\ freedom\ for\ loc * geno(block)}\right)\right]}{\text{rl}}$$

Where rl is the Expected Means Squares coefficient for Genotype.

Genetic correlations were then determined for each of the 28 pairs of traits, using the genetic covariance and genetic variance components for each trait. Correlations were calculated using the method proposed by Johnson, Robinson, and Comstock (1955a & 1955b). Standard errors for each correlation coefficient were calculated using Mode and Robinson's (1959)

method and standard errors for genetic covariances were calculated from their equation. The genetic coefficient of variance was also calculated for each trait by taking the square root of the genotypic variance component and dividing by the mean for each trait.

Heritabilities were estimated on an entry mean basis by the method described in Johnson et al (1955a). Confidence intervals (95%) for each trait's heritability estimate were calculated using the Knapp, Stroup, and Ross (1985) method. SAS was used to obtain the appropriate F values in the equation and are listed below.

For C.I of H RSII: F-value upper=0.75635, F-value lower=1.30307, df1=110, df2=223

For C.I of H RSIII: F-value upper=0.74759, F-value lower=1.31710, df1=102, df2=204

Phenotypic Covariances were calculated using the Genotype cross product elements of the Hypothesis Type III SSCP Matrix from the MANOVA analysis in SAS under PROC GLM in the following manner (SAS Institute 2003):

$$\left(\frac{H=Type\ III\ CP\ Matric\ for\ each\ pair/df}{rl}\right)$$

Where rl is the Expected Means Squares coefficient for Genotype.

Phenotypic correlations were then determined for each of the 28 pairs of traits, using the phenotypic covariance and phenotypic variance components for each trait. Correlations were calculated using the method proposed by Johnson et. al. (1955a & 1955b). Standard errors for each correlation coefficient were calculated using Mode and Robinson's (1959)

method and standard errors for phenotypic covariances were calculated from within their equation. Means for each of the eight traits was calculated in the SAS system using the lsmeans statement in PROC MIXED, since mixed model handles missing values better, with Genotype being fixed in the model statement (SAS Institute 2003). A Least Significant Difference (LSD) was calculated for each trait using the LSD equation:

$t_{\alpha/2} *(\sqrt{(2*MS_{GL}/n)}$

where: $t_{\alpha/2} = 1.98$ for $\alpha=0.05$ and df = (RSII) 110 and (RSIII) 96 and n = 6.

Contrasts were used to test the association of pubescence color with yield as well as flower color with yield using the 'contrast' statement in the Proc GLM procedure in SAS to determine any association between the two traits. Yield mean rankings for each location were evaluated using PROC CORR in the SAS system, with the Spearmen Correlation Coefficient to evaluate genotype and environment interactions (SAS Institute 2003). Allele frequencies were also determined for each population for both flower color and pubescence color with the following equation: Frequeny(A) = Frequency of homozygous (AA) + ½ the frequency of heterozygotes (Aa).

Theoretical Expected Gain from an imposed selection experiment on yield, oil, and protein were calculated using the method from Falconer and Mackay (1996). ($Rx=i\sigma_{px}H_x$, where R is the response to selection, i is the intensity of selection, σ_{px} is the square root of the

phenotypic variance, and H_x is the broad sense heritability). The indirect selection response for yield, oil, and protein from each of the directly selected traits was calculated using the method described by Falconer and Mackay (1996). ($CR_y=i\sigma p_y\sqrt{H_y}\sqrt{H_x}r_g$, where CR_y is the indirect response to selection on trait y with direct selection on x, i is the intensity of selection, σ_{py} is the square root of the phenotypic variance of trait y, and $\sqrt{H_x}$ is the square root of the broad sense heritability of trait x, and $\sqrt{H_y}$ is the square root of the broad sense heritability of trait y, and r_g is the genetic correlation between x and y). Selection intensity values were obtained from Becker (1984) with selection for the top ten percent of each population.

Results and Discussion

Population Means

These populations were developed in the early seventies from breeding material and cultivars of that time (Appendix A1 & A2) with the exception of the unadapted plant introduction germplasm of the parents in RSII. So it is of interest to compare the productivity of modern cultivars to the productivity of lines in RSII and RSIII to determine if any genetic progress had occurred towards increasing yields in the absence of artificial selection.

RSIII had slightly higher means and the mean ranges than RSII for all traits measured but with some overlap in values (Tables 2 & 3). The mean yield of RSII was 2070 ± 337 kg/ha and the mean yield of RSIII was 2205 ± 252 kg/ha. The mean yield ranges in RSIII and RSII were similar with about 1400 kg/ha between the high and low values. Flowering date, maturity date, and lodging all have similar high and low values between the two populations. The lowest calculated mean value for percent protein (39.3%) was about the same in both populations while the highest calculated value for percent oil (20.7%) was about the same in both populations. Seed weight in RSIII had a wide range with the highest genotype mean of 19.9 g and the lowest of 11.2 g.

NC-Roy and NC-Raleigh were grown as check cultivars in the RSIII test. NC-Roy yielded an average 3306 kg/ha, which was 5% higher than the best yielding genotype in maturity group

VI which yielded 3142 kg/ha. This genotype (N-412, maturity group VI, Appendix B4) was also the highest yielding line tested in RSIII. NC-Raleigh outperformed all maturity group VII genotypes in RSIII, with an average 3003 kg/ha yield, which was about 2% higher than the highest yielding group VII genotype in RSIII which yielded 2956 kg/ha. These differences were not statistically significant however.

NC-Raleigh was grown as a check cultivar in the RSII test and did not yield as high when compared to its performance in RSIII, with an average 2745 kg/ha yield. This was 3% lower than the best yielding group VII genotype in RSII, with a yield was 2825 kg/ha. This genotype (N-239, maturity group VII, Appendix B3) was also the highest yielding line tested in RSII. Of the maturity group VI genotypes in RSII, the best yielding genotype had an average yield of 2600 kg/ha. Although there were yield differences in the checks across populations, visual differences between growing conditions were not observed.

According to the 2009 North Carolina Official Variety Test (NCOVT), the top yielding cultivars and experimental lines in maturity group VII-VIII yielded between 3091 kg/ha and 3366 kg/ha. Of those cultivars tested in group VII-VIII, NC-Raleigh, a group VII cultivar, yielded 2694 kg/ha, 10% less than when grown in RSIII and 2% less than when grown in RSII. In maturity group VI, the top yielding cultivars and experimental lines yielded between 2916 kg/ha and 3507 kg/ha. Of those cultivars tested in group VI, NC-

Roy yielded 3050 kg/ha, 8% less than when grown in RSIII. (Bowman, 2009a & 2009b) The 2008 statewide North Carolina soybean yield average was 2217 kg/ha.

By comparing the yields of the NCOVT to the means of the two populations in the current study, it may lead to determining if lines could be developed through random mating that compared well to those under artificial selection, this was not tested for significance however. When comparing the yield of those genotypes in RSII and RSIII to the NCOVT and NC statewide average, RSIII stayed competitive with the NCOVT and performed slightly better than RSII. RSIII had one genotype which yielded as much as those lines tested in the NCOVT while RSII fell short 7.5% with its highest yielding genotype. This might be expected given the comparison of the cultivar NC-Raleigh's performance between RSII, RSIII, and the NCOVT. NC-Raleigh's yields are similar for NCOVT and RSII, but when grown under the same conditions as those genotypes in RSIII, NC-Raleigh's yields increased around 10%. These differences described are not large and are within the LSD ranges of RSII and RSIII, of 620 kg/ha and 528 kg/ha respectively. So while there are noticeable differences, they may not be statistically significant.

Suneson (1956) proposed an evolutionary plant breeding method to improve selection and testing efficiency. Crosses were made and the populations were exposed to natural selection for improvement. Composite Cross XIV was developed from nine adapted barley varieties

which included a male sterile parent. He concluded that natural selection allowed for progressive yield improvement in the composite crosses. Suneson suggested that 15 generations of random mating was desirable for yield improvement and continued through natural selection, cyclic hybrid recombinations with natural selection, or direct artificial selection. Soliman and Allard (1991) reviewed the work of Suneson and others that studied composite crosses of barely. They determined that any yield improvement among composite crosses was due to natural selection.

It appears that random mating with the possible aid of natural selection also provided yield gains in RSII and RSIII. Considering the LSD for each population was over 500 kg/ha, most NCOVT lines and those check cultivars grown with the populations did not greatly exceed some of the top performing genotypes tested. Therefore, it is likely that with random mating these two populations stayed competitive with current varieties.

Population Variation

Through the ANOVA table, (Tables 4 & 5) significant genotypic variability for all traits was found and location by genotype interactions were determined significant for all traits. Genotypic variances were determined and were significant for each trait in each population (Tables 6 & 7).

These two populations were grown contiguously in the same field but tested in different experiments, therefore preventing direct statistical comparison. However, the expectation that these populations are genetically similar to an extent may allow inferences on how they differ. Since the means, ranges, and genetic variances for maturity and flowering date are very similar between both populations, comparisons for other traits such as yield, percent protein, and percent oil can be made. The genotypic variance for yield in RSIII nearly doubled that of RSII. This was observed in seed weight as well. The genotypic variance for percent oil in RSIII was four times that in RSII and for percent protein RSIII genotypic variance was three times that in RSII. Plant height was the only trait where genotypic variance was greater in RSII than RSIII. These differences cannot be tested statistically, so while estimates in RSIII appear larger than RSII, the differences may not be significant.

A method of comparison is to evaluate the relative magnitudes of the genetic coefficient of variation, the GCV (Tables 6 & 7). Similar GCV values were obtained for maturity date in both populations while flowering date, lodging, and height were higher in RSII than RSIII. Yield, seed weight, percent protein, and percent oil had higher GCV values in RSIII than in RSII, with seed weight and percent oil being twice as large. This indicates that larger genetic variation was present in RSIII than RSII.

With significant genotype by environment interactions for each trait in each population, except for lodging in RSIII, yield mean rankings by location were compared to examine the interactions more closely (Table 8). Rank correlation coefficients for each pair of locations were calculated. Small positive correlations were noted in both populations between the Clinton and Plymouth locations. A small positive correlation was also significant in RSIII between Plymouth and Mount Olive. This suggests that the environments in Clinton and Plymouth were similar with fewer significant genotypic interactions across those two locations. As well, the Mount Olive and Plymouth locations in RSIII were similar. No correlation was observed in RSII between Mount Olive and Clinton or Plymouth and no correlation in RSIII was observed between Mount Olive and Clinton. These similarities and differences between locations are likely due to moisture stress and minor infestations of phytophthora root rot in RSII at Plymouth.

Another method of evaluating the genotype by environment interactions is to compare the top and bottom 20% yielding genotypes across locations to determine if any genotypes that are in the top 20% at one location are also in the bottom 20% in another location (Table 9). Evidence of this was observed in 11 genotypes in RSII and 13 in RSIII. The top performing RSII genotype at Plymouth was also in the bottom 20% at both Mount Olive and Clinton. The highest ranking genotype in RSII at Mount Olive was also one of the genotypes

found in the bottom 20% at Clinton. RSIII had three top ten genotypes at one location also rank in the bottom 20% at one or both of the other locations. This may indicate that the environments at each location differed enough relative to each genotype to produce differences in their genetic potential. Since many lines changed ranks between locations, more environments are needed to obtain an accurate estimate a of genotype's mean potential.

Heritability

Heritabilities on an entry mean basis and their confidence intervals were calculated for each trait (Tables 6 & 7). As expected, given the lower genetic variance and higher genotype by environmental interaction estimates, RSIII's heritability estimates were generally higher than those for RSII.

Lodging and plant height were the exceptions with higher heritability estimates in RSII. RSII had an estimated yield heritability of 33.1% but had a very wide confidence interval. The yield heritability estimate in RSIII was higher than RSII at 55% with a much narrower confidence interval. Similar heritabilities between the two populations were observed in flowering date, 94%, plant height, 82%, and maturity date, 90%, with similar confidence intervals. The heritability estimate for lodging in RSIII was 78% while RSII was slightly higher at 85%, both with narrow confidence intervals. Seed weight in RSIII had the highest heritability estimate of 95%, and also had the narrowest confidence interval. Seed weight

heritability in RSII was much lower than RSIII at 73% and also had a fairly narrow confidence interval. Percent protein and percent oil had very high heritability estimates in RSIII at 87% and 91% respectively, both with narrow confidence intervals. Lower heritability estimates for the two traits in RSII were observed, 68% for percent protein and 56% for percent oil.

Comparing the heritability estimates of these populations to estimates reported in the literature can provide a means of assessing random mating's influence on a population. Burton (1987) summarized heritabilities on an entry mean basis that were calculated from eight biparental populations for nine traits and compared them. Seed yield had the lowest heritability (ranged from 3% to 58%). Higher estimates were reported for height (ranged 70% to 90%), maturity (ranged 75% to 94%), and protein (ranged 57% to 90%). Lodging, seed weight, and oil also had higher heritabilities than seed yield but with wider ranges (43% to 75%, 44% to 94%, and 51% to 89% respectively). The study by Burton et. al. (1990) which evaluated these same populations after four cycles of yield recurrent selection found yield heritability estimates as high as 62% in cycle 1 to as low as 43% in cycle 4 for RSII. In RSIII they found yield heritability estimates of as low as 18% in cycle 1 to as high as 32% in cycle 3.

The yield heritability estimate in RSII was similar to the cycle four estimate in the study by Burton et. al. (1990) and RSIII had a much larger yield heritability estimate than those found by Burton et. al.(1990). Therefore, little progress on yield has occurred in RSII while in RSIII, with an increased yield heritability, progress may have been made through natural selection. Estimates of the other seven traits in the two populations in this study were within or above those estimates presented by Burton (1987) with percent oil in RSIII being well above those estimates summarized. Percent protein in RSIII was within the ranges provided by Burton (1987). This indicates that expected gain from selection on these traits in RSII and RSIII would be greater than the expected gain in the studies presented by Burton (1987). These results indicate that random mating and natural selection can be an efficient method to increase heritability which would result in more gain from selection.

Population Correlations

Genotypic and phenotypic correlations along with their standard errors were obtained for each population for each set of 8 traits, 28 total comparisons (Tables 10 & 11). Genotypic and phenotypic correlations were considered to be significant if their standard errors were less than half of the calculated correlation coefficient. Based on this criterion, the positive genotypic correlations in RSII (Table 10) included maturity with flowering date (0.45), flowering date with lodging (0.3), flowering with seed weight (0.11), flowering date with plant height (0.42), maturity date with plant height (0.23), lodging with plant height (0.43),

maturity date with seed weight (0.43), maturity date with yield (0.29), lodging with percent oil (0.24), and height with seed weight (0.22). Negative genotypic correlations were found between flowering date and percent protein (-0.15), flowering date with yield (-0.23), lodging and yield (-0.39) and between percent protein and percent oil (-0.61). In RSIII (Table 11) positive genotypic correlations included maturity with flowering date (0.43), flowering date with lodging (0.35), flowering date with plant height (0.45), lodging with plant height (0.29), maturity date with seed weight (0.28), and seed weight with yield (0.63). Negative genotypic correlations in RSIII were found between flowering date and yield (-0.43), flowering and percent protein (-0.22), lodging and seed weight (-0.27), lodging and yield (-0.6), and percent protein and percent oil (-0.73). Similar phenotypic correlations were obtained in both RSII and RSIII. In RSII, seed weight and yield were positively phenotypically correlated (0.24) in addition to those genotypic correlations. The only additional phenotypic correlation that was not in the genotypic correlations in RSIII was the negative correlation between percent protein and maturity (-0.19).

The genotypic and phenotypic correlations between percent protein with percent oil were negative, with an absolute value greater than 0.6 in both populations. This is consistent with published literature, showing a strong negative correlation between the two traits. Burton (1991) had correlations greater than 0.6 in all three populations studied, two of which were randomly intermated for 5 generations. Johnson et. al. (1955b) observed a negative

correlation with an absolute value of 0.48. Negative correlations between yield and percent protein tend to be significantly negative in literature as well. Burton (1991) reported an estimate as high as -0.47 and Johnson et. al. (1955b) reported an estimate of -0.64. But in RSII, that correlation was negative but not significant and in RSIII the estimate was neither significant nor negative. This is important for breeders since in these populations selection for yield or protein will not negatively impact the other. Also since this was achieved in this population, perhaps it can be achieved in other populations with the right breeding method (one that decreases LD, recurrent selection).

Burton (1987) reviewed correlation coefficients between traits in many populations and revealed that yield and percent oil tend to have a low correlation. This was also observed in the present study. RSII had a non significant correlation of 0.07 and RSIII had a correlation of 0.26, which was found to be non significant as well. So selection on yield or oil will increase the other, and with a high heritability estimate for oil in both populations, selection for oil might be favored to increase both traits together. Yield and lodging tend to be negatively correlated in breeding populations, and this was also true of the two populations in this study.

Genetic material used by Johnson et. al. (1955a) involved two populations derived from the F_3 of biparental origins. The populations evaluated by Burton (1987) were also of biparental

origins. The two populations evaluated in the current study are inbred lines derived from many generations of random mating, developed from multiple parents. Therefore, linkage phase disequilibrium has been greatly reduced in RSII and RSIII. With high levels of genetic recombination, LD would be reduced, making these estimates and correlations between the traits less affected by linkage, except for those loci that are very closely linked (Appendix C). Also, the populations studied in Burton (1987) and Johnson et. al. (1955b) had less inbreeding than materials in the current study and therefore dominance may have affected their estimates. While the two populations of this study involved inbred line therefore the estimates in these population were unaffected by dominance.

Tinuis et. al. (1993) evaluated three subpopulations of RSIII in which nine cycles of random mating were conducted followed by three cycles of recurrent mass selection. The combined phenotypic correlation between protein and oil was -0.89, while the phenotypic correlation between yield and protein was -0.33, which was found to be not significant. Yield and oil phenotypic correlation was non significant by positive at 0.47. The RSIII population after 26 generations of random mating had the same result between yield and oil and between oil and protein, although the current RSIII population had a lower correlation. The correlation between yield and protein is no longer negative but still non significant in the current RSIII population. This indicates that with random mating, these correlations decreased, most likely a result of more genetic recombination events reducing linkage phase disequilibrium.

Breeders can use these findings to implement new breeding strategies during selection. Selection on yield in these populations may be an efficient way to indirectly increase percent oil and at the same time keep percent protein constant. Selection on percent oil in these populations will still have a negative impact on percent protein, and vice versa. Depending on whether correlations are due to gene linkages, pleiotropy, or the physiological processes of the plant is of great interest to breeders and has been debated in their studies. Johnson et. al. (1955b) did not rule out genetic linkages in their study but also suggested that the underlying cause was likely due to the plant's physiological process for those correlations with larger coefficients. Burton (1991) suggested that observed protein, oil, and yield correlations in the two populations with random intermating were likely due to physiological processes and not genetic linkages. Therefore, using a breeding method to increase genetic recombination would increase the likelihood of accumulating favorable genetic combinations. Selection during the breeding method would also keep those favorable combinations in the population.

Likely explanations can be found for some of the genetic correlations observed in these populations Flowering date and maturity date may be correlated in soybeans since each stage of development is dependent on the previous one. Another association in soybeans is height with lodging. It is possible that as plant height increases lodging may also increase given the additional height for the plant to support. Seed weight and lodging may be positively

correlated as well since as the seed weight increases, the plant has more to support, likely causing it to lodge. The positive correlation between percent oil and maturity can be explained by Wolf, Cavins, Kleiman, and Black (1982) that showed that oil content is positively correlated with temperature. Therefore, as the growing season continues, the temperature falls, and oil content decreases, giving earlier maturity groups higher percent oil. Seed size (seed weight of 100 seeds) and yield are often positively correlated as well. If seed size is large then the total yield will increase if seed number stays constant since yield is equal to the product of seed number and seed size.

If pleiotropy is acting on the negative correlation between percent protein and percent oil then having a high protein and high oil phenotype would not be possible. Genotypic results of the two populations in this study support that idea, since with genes assorting independently, gene combinations that would produce both high percent protein and high percent oil were not observed in either population. In RSII the individual with the highest mean protein of 43.68% had a percent oil mean of 18.7% with a yield of 2312 kg/ha. The individual with the highest mean percent oil, 20.7% had a percent protein mean of 40.9% with a yield of 1656 kg/ha. The individual with high means in both traits had 42.17% protein and 19.82% oil with a yield of 1673 kg/ha. In RSIII the individual with the highest mean protein of 45.06% had a percent oil mean of 15.85% with a yield of 2035 kg/ha. The individual with the highest mean percent oil, 20.99% had a percent protein mean of 39.52%

with a yield of 1964 kg/ha. The individual with high means in both traits had 42.33% protein and 19.73% oil with a yield of 2170 kg/ha. Those genotypes that have a high value for protein also have low values for oil and vise versa, indicating that gene combinations favoring high values of both traits did not occur. Yield in these genotypes did not differ all that greatly however but given the low correlations between yield and protein and yield with oil, yield would not be dependent on the other values.

Gene Frequencies

A major difference in a random mating population and a biparental, artificially selected population is expected gene frequencies. With a biparental population, initial gene frequencies for a single locus with two alleles is $p = q = 0.5$. In random mating populations with multiple parents, gene frequencies can involve multiple alleles, making calculations more difficult. For a simple two allele, single locus example, pubescence color and flower color gene frequencies were estimated in the initial population as well as in the current population. Since 50% of the genetic makeup in both populations was provided by N69-2774, 50% of the gene frequencies will consist of white flowers with gray pubescence. The other 50% will involve a combination of the other parents in the pedigree. In RSIII, the expected initial gene frequencies for flower color was 75% white, 25% purple. Pubescence color was 67% gray, 33% tawny. The random mating progeny of RSIII had flower color gene frequencies of 49% white and 51% purple and pubescence color gene frequencies of 58%

tawny and 42% gray. In RSII, the expected initial gene frequencies for flower color was 58% white, 42% purple. Pubescence color was 60% gray and 40% tawny. The progeny after 26 generations of random mating in RSII had flower color gene frequencies of 53% white and 47% purple and pubescence color gene frequencies of 80% tawny and 20% gray (Table 12). Taking sample size into account, the variance in change of gene frequencies can be expected to be small enough to disregard, having a 10^{-4} value. The expectation that gene frequencies will remain unchanged in a random mating population is not demonstrated with these populations. A Chi-squared test rejected the hypothesis that gene frequencies remained constant in these two populations except for flower color in RSII (Table 12).

One possible explanation for deviation from expected is that the number of breeding individuals did not equal the number of progeny in each generation. At the end of each random mating cycle, approximately 200 male sterile plants were harvested and equal number of seed was kept from each male sterile plant and planted the following year. But, this was not constant every year. Thus, if the number of progeny from each parent in each generation was not the same, a change in gene frequencies would likely be observed. If any other expectation of an ideal population were not met, a change in gene frequencies might also be observed. For example, if random mating in RSII and RSIII was not completely random, gene frequencies would change over time. In a study by Graef, Locke, Specht, and Lee (1991), they observed unexpected gene frequencies after only seven cycles of random

mating. They believed that the random mating in their population was not completely random. Differences in time of flowering and the effects of natural selection were two explanations they cited. The same two explanations could apply to RSII and RSIII, which could explain some of the deviation from the initial gene frequencies.

Farmers and breeders tend to have preferences when it comes to the soybean's agronomic appearance. It is known that many farmers tend to prefer tawny pubescence over gray and most cultivars released by major breeding companies tend to have tawny pubescence. So it is also of interest to discern a possible connection between pubescence color and yield. Given the many generations of random mating, there has been more genetic recombination in these populations than is usually observed in biparental inbred lines. Thus, with these populations, pleiotropic effects between yield and pubescence color can be determined. The same can be determined for flower color and yield although no preference is usually given for purple or white. In both populations no statistically significant association was found between yield and pubescence color or in yield and flower color. Therefore, tawny pubescent cultivars do not tend to have higher yields.

Theoretical Response to a Selection Experiment

In the current study, theoretical expected gain from both direct and indirect selection in these two populations of inbred lines was calculated for yield, percent protein, and percent oil

(Table 13). A selection intensity of 10% for 120 lines in RSII and 100 lines for RSIII was used in each population giving 11 selected in RSII and 10 selected and in RSIII. Expected gain in yield in RSIII was estimated at 272 kg/ha while expected yield gain in RSII was only 161 kg/ha. Percent protein and percent oil demonstrated the same trend of greater gain in RSIII compared to RSII. Expected gain in percent protein was estimated at 1.96% and gain in percent oil was expected to be 1.47% in RSIII while in RSII expected gain in percent protein was only 1.0% and expected gain in percent oil was 0.57%.

Considering the correlated response of percent protein and percent oil to direct selection on yield in RSIII, there was a slight gain in percent oil, 0.29% and a very slight change in percent protein, 0.11%. The correlated response in RSII of percent oil to selection on yield gave similar results as RSIII, with an expected gain of 0.03% while percent protein had a decrease, -0.12%. Given the small correlations of yield with percent protein and percent oil in both populations, it would be expected that yield would increase with only slight changes in oil and protein. When the response of indirect selection on percent protein and yield to direct selection on percent oil was calculated, RSIII had an increase in yield of 90 kg/ha and had a decrease in percent protein of -1.46%. In RSII, an expected gain of 14.42 kg/ha is expected on yield and a decrease of -0.56% is expected on percent protein with direct selection on percent oil. Therefore with direct selection on oil, it is expected that percent oil and yield would increase in both populations. RSIII had a greater decrease in percent protein

when percent oil was being selected when compared to RSII, so a greater loss in protein would be expected in RSIII. When considering the correlated response of percent oil and yield to direct selection on percent protein, an increase in yield of 23.17 kg/ha and a decrease in percent oil of -1.05% is predicted in RSIII. RSII on the other hand is expected to have a decrease in both traits with selection on percent protein. Yield is expected to decrease 39.551 kg/ha and oil is expected to decrease -0.38% in RSII. Again, RSII's correlated response to selection on percent protein had a smaller deleterious effect on percent oil than in RSIII, with a yield loss in RSII and not in RSIII.

Expected gain in direct and indirect selection in the current populations do not agree with the experiments using these same populations but in earlier generations of random mating and inbreeding. Burton et. al. (1990) reported greater yield gains in RSII compared to RSIII, while in the current study, a greater expected gain in yield in RSIII will be more likely than RSII. But, the correlated responses on percent protein and percent oil were similar between the current study and populations in Burton et. al. (1990). While Burton et. al. (1990) found greater genetic variability in RSII, RSIII in the current population has greater genetic variability, accounting for the differences in the two studies.

One possible explanation for the differences in genetic variability in Burton et. al. (1990) and the current study may be from a genetic bottleneck in RSII as random mating continued.

During the development of RSII, only 10 male parents of the highest yielding lines from a recurrent selection experiment were used to form the random mating population. Therefore, a genetic bottleneck in the selection of those10 male parents may have greatly decreased the amount of genetic variability in RSII compared to that of RSIII which had no selection during population development. Another concern with the development of RSII is the large input of genetic material from D49-2491. Since D49-2491 contributes 75% of the genetics in the male pollen during pollination of the male sterile maintainer line, N67-2774, only 25% of the nine plant introductions is represented with each allele only being present with 1.4% frequency. This low gene frequency of the nine plant introduction subjects their genes to complete loss due to random genetic drift. Drift is likely to cause a complete loss of genes that appear in low frequencies if they do not contribute advantages. As random mating in RSII continued past the generations studied by Burton et. al. (1990), no new genetic variability was added and due to drift the similar population tended towards fixation, diminishing genetic variability even more. Equal genetic contribution of all the male parents in RSIII should lead to higher genetic variability in the progeny since random genetic drift would have less of an impact and a lower rate of gene loss might be expected given higher frequencies of genes from the male parents.

Need for Recombination

Stefaniak, Hyten, Pantalone, Klarer, and Pfeiffer (2006) state that recombination will have great impacts on polygenic traits, such as seed yield. If seed yield has an additive gene model, recombination combines favorable alleles together to optimize yield. Therefore, any breeding method that allows for increased recombination should be implemented. The authors also suggest that during natural selection, alleles are favored during gametophyte development, as only one-fourth of the gametes survive in the female. They also determined that in a biparental cross between Williams and Essex that any advancement in selection was due to an additive gene model. Once favorable gene combinations were established, low recombination rates would be desired to keep those favorable combinations intact.

Recombination hotspots have been found to be located mainly in genic regions of the chromosome compared to the regions between genes. Additive gene models as well as epistatic interactions will benefit from the high levels of recombination combining favorable alleles. This may be noted in RSII and RSIII when evaluating yield. Comparing parental data to progeny, yield in RSII has improved greatly. Favorable alleles in the nine plant introductions were able to combine with the alleles of D49-2491 and N69-2774 to produce higher yielding progeny, even at their low frequency. Random genetic drift would not cause a loss if those alleles are favorable. Although a direct comparison may not be made the observation that the highest yielding parent in RSII reaching a yield of only 2210 kg/ha,

having a genotype in RSII with a mean yield of 2825 kg/ha can be attributed to recombination and natural selection combining and keeping those favorable alleles. The same is true in RSIII, which has more equal frequencies of parental alleles than in RSII. With the highest yielding parent in RSIII having a yield of 2668 kg/ha, the occurrence of a genotype in RSIII with a mean yield of 3142 kg/ha can also be attributed to favorable recombination, but again a direct comparison cannot be made.

Conclusion

These two populations were randomly intermated for 26 generations without artificial selection but were subjected to natural selection in each generation. Random mating and natural selection are rare methodologies used in breeding for self-pollinated species making this study unique. Yield means in these populations compared well to check cultivars and those elite cultivars being tested in the NCOVT, with a few genotypes in these two populations meeting their yields. Heritabilities in these two populations were high, with some traits having an estimated value over 90%. Genetic correlations were estimated and less than half of the traits tested were correlated with another trait, some with very little statistical significance.

Negative linkage phase disequilibrium occurs in populations by their development, usually of elite by elite crosses. RSII was likely in negative linkage disequilibrium phase, according to Sileva and Diez-Barra (1985), and given the parents of RSIII were well adapted selected cultivars, recurrent selection would prove most beneficial for this population. With the predicted direct and indirect selection responses calculated for RSIII, much progress can be expected from selection in that population. There is possible gain in yield, percent protein, and percent oil given the right selection intensity and selection method. Kenworthy (1979), evaluated different crossing schemes to identify the benefits of each crossing method when using exotic germplasm. Two way crosses were believed to be most beneficial when trying to

breed for pest resistance. Three way crosses were believed to be the most beneficial when trying to breed for increased yield. Intermating however proved to be the most beneficial for increasing genetic variance, which is needed to make any improvement during selection. Considering RSII is partially genetically composed of nine unadapted plant introductions, intermating or recurrent selection should be the most beneficial breeding scheme to achieve improvement during selection. To prevent the loss of the exotic germplasm through random genetic drift, due to their low gene frequeinces in the population, increasing the frequency of those exotic lines by starting with more plants of that germplasm may be a solution.

The uses of genetic male sterility are potentially endless when considering Kenworthy's (1979) suggestions on bringing in exotic germplasm. He stated that with recurrent selection, increased genetic variability is expected. With greater genetic variability, greater improvement can be expected during selection. He also brought to attention the limited genetic base from which most soybean varieties have been developed. With genetic male sterility and recurrent selection, bringing in new genetic variability can greatly enhance the potential improvement in soybeans.

Possible limitations of the current experiments might include too few genotypes being tested as well as inadequate location and year testing. Evaluating these two populations under ideal experimental conditions is important. By sampling only 105-120 genotypes and testing in

only three environments, the estimates calculated may not be robust. Hartl and Clark (2007) consider a population to be small if it is under 500 individuals. Therefore, the 100 genotypes might be considered too few for efficient testing. Also, more environments over multiple years will result in more accurate estimates of population parameters and line performance. With only one year of data, genotypic variance will be inflated. Having more genotypes per population can also be expected to result in more accurate estimates. Problems arise from having more genotypes to test, as space and time can be limited to accurately measure and test each genotype as well as field variability within a test.

The genotypes in these populations can also be utilized in biparental breeding methods. Since $F_{5:8}$ seed was harvested from each genotype, relatively homozygous lines are available. Crosses can be made between these homozygous genotypes and current cultivars, both with desired traits but contrasting phenotypes to achieve higher genetic variance among the progeny. Selection will also increase the frequency of those favorable gene combinations since selection would keep them in the population. With the increased gene recombination, more opportunities occurred to combine favorable genes.

These populations can also be evaluated to test for the amount of genetic variability left in the populations. This would be of interest to determine what effect random mating has

on a closed population. Selection for yield, percent protein, and percent oil can be carried out separately in a recurrent selection breeding method. Once a plateau has been thought to be reached, selection should be started in the opposite direction. If the genes are fixed, no phenotype changes will be observed, if selection decreases the phenotype of the trait, fixation has not occurred, meaning genetic variability persists in the random mating population with selection. This will allow determining the possible influence the nine plant introduction in RSII had on the progeny. If any genetic variably persists, it will most likely be caused from the initial low frequency of those alleles being favored by natural selection and therefore not lost.

Yield trials will be performed in the summer of 2010 with those top 10% yielding genotypes from each population. Those available parents from each population will be grown in the experiment as well. Four locations will be used with current cultivars as check varieties. This will allow for a more accurate comparison of these populations with their parents and current cultivars to determine if yield has increased during the 26 generations of random mating without artificial selection.

If a random mating experiment like this were to be performed again, choosing the proper parents to develop the population would be important. To allow for maximum genetic variability, parents would include cultivars with distinct differences in percent oil, percent

protein, and yield with both high and low values of all three traits being represented. These parents would be crossed with a genetic male sterile maintainer line. Then, the F1 progeny would be backcrossed to the male parents to increase their gene frequencies. Those BCF1 plants would then be placed in an open pollinating crossing block to allow for genetic recombination among the lines. 15 generations of random mating without selection would be suggested, based on Suneson (1956) results. After the15th generation, recurrent selection would be implemented in the population. Selection for yield, percent protein, and percent oil could be carried out during each recurrent selection cycle, with genotypes being selected based on one, two or all three traits. With selection in each generation, new gene combinations formed through random mating will be selected for and therefore carried into the next generation. This will allow for only favorable gene combinations being carried to the next generation, since random mating breaks both favorable and unfavorable gene combinations. To increase genetic variability in each generation, new parental genotypes could be added to the population every 5 generations of random mating. This will ensure genetic variability through this long term recurrent selection experiment. But, by bringing in new parental genotypes, those alleles will be subjected to random genetic drift since their alleles would be at such a small frequency in the population. To prevent that, these new alleles need to be favored in the selection procedure. Also by bringing in new parental genotypes in at higher frequencies random genetic drift may not be a factor. Future breeding procedures can benefit greatly by utilizing genetic male sterility in a recurrent selection

program. With proper parent selection that maximizes genetic variability, improvement can be expected in any population.

REFERENCES

Anderson R. L. & Bancroft T. A. (1952). *Statistical Theory in Research. New York, NY: McGraw-Hill.*

Ash, M., & Dohlman, E. (26 Mar 2008). *USDA Soybean Projections, 2008-17.* Retrieved from: http://www.ers.usda.gov/briefing/soybeansoilcrops/2008baseline.htm

Ash, M., Dohlman, E., & Wittenberger, K. (4 Dec 2008) *Soybeans and Oil Crops: Background.* Retrieved from: http://www.ers.usda.gov/Briefing/SoybeansOilcrops/background.htm

(27 Aug 2009). *Average Protein and Oil at 13 percent Moisture.* Retrieved from: http://www.unitedsoybean.org/programs/soy_measurements/2008/average_protein_and_oil_at_13_percent_moisture.aspx

Bahrenfus J. B., & Fehr W. R. (1980). *Registration of Vinton Soybean. Crop Science, 20,* 673a-674b.

Becker, W. A. (1984). *Manual of Quantitative Genetics. (4th ed.) Pullman, Washington: Academic Enterprise.* 147-157.

Brim C. A., & Young M. F. (1972). *Registration of a Male-Sterile Maintainer Line (N69-2774 of Soybeans (Reg. No. GP 12). Crop Science, 12,* 399b.

Brim C. A., & Burton J. W. (1979). *Recurrent Selection in Soybeans. II. Selection for Increased Percent Protein in Seeds. Crop Science, 19,* 494-498.

Boerma, H.R. (1979). *Comparison of Past and Recently Developed Soybean Cultivars in Maturity Groups VI, VII, and VIII. Crop Science, 19,* 611-613.

Bowman, D. T. (2009a). *Table 4: Data Combined Over Locations For Conventional Group VI Non Stressed Soybean at North Carolina – 2009. North Carolina Measured Crop Performance Cotton and Soybean 2009.*

Bowman, D. T. (2009b). *Table 6: Data Combined Over Locations For Conventional Group VII-VIII Non Stressed Soybean at North Carolina – 2009. North Carolina Measured Crop Performance Cotton and Soybean 2009.*

Burton, J. W., & Brim, C. A. (1981). *Registration of Two Soybean Germplasm Populations. Crop Science, 21*, 801.

Burton, J.W. (1987). Quantitative Genetics: Results Relevant to Soybean Breeding. In J.R. Wilcox (2nd Ed.) *Soybeans: Improvement, Production, and Uses* (pp. 211-247). Madison, WI: American Society of Agronomy.

Burton, J. W., Koinange, E.M. K., & Brim, C. A. (1990). *Recurrent Selfed Progeny Selection for Yield in Soybean Using Genetic Male Sterility. Crop Science, 30,* 1222-1226.

Burton, J. W. (1991). Development of High-Yielding High-Protein Soybean Germplasm. In R. F. Wilson. *Designing Value Added Soybean for Markets of the Future* (pp. 109-117). Champaign, IL: American Oil Chemists' Society.

Burton, J. W., Carter, T. E. Jr., & Wilson, R. F. (1999). *Registration of 'Prolina' Soybean. Crop Science, 39,* 294a-295a.

Falconer, D. S. & Mackey, T. F. C. (1996). *Introduction to Quantitative Genetics* (4th Ed.). Harlow, England: Pearson Education Limited.

Gizlice, Z., Carter, T. E. Jr., & Burton, J. W. (1994). *Genetic Base for North American Public Soybean Cultivars Released between 1947 and 1988. Crop Science, 34,* 1134-1151.

Greaf G. L., Locke, C. R., Specht, J. E., & Lee, D. J. (1991). Allele Frequency Changes During Seven Cycles of Forced Random Outcrossing in Soybean. *Agronomy Abstracts* (pp. 95). Madison, WI: American Society of Agronomy.

Hartl, D., & Clark, A. G. (2007). *Principles of Population Genetics* (4th Ed.). Sunderland, MA: Sinauer Associates, Inc.

Helms, T. C., & Orf, J. H. (1998). *Protein, Oil, and Yield of Soybean lines Selected for Increased Protein. Crop Science, 38,* 707-711.

Johnson, H. W., Robinson, H. F., & Comstock, R. E. (1955a). *Estimates of Genetic and Environmental Variability in Soybeans. Agronomy Journal, 47,* 314-318.

Johnson, H. W., Robinson, H. F., & Comstock, R. E. (1955b). *Genotypic and Phenotypic Correlations in Soybeans and Their Implications in Selection. Agronomy Journal, 47,* 477-483.

Kenworthy, W. J. (1979). Strategies for Introgressing Exotic Germplasm in Breeding Programs. In F. T. Corbin, *World Soybean Research Conference II: Proceedings* (pp. 217-223). Boulder, CO: Westview Press.

Knapp, S. J., Stroup, W. W., & Ross, W. M. (1985). *Exact Confidence Intervals for Heritability on a Progeny Mean Basis. Crop Science, 25,* 192-194.

Mode, C.J., & Robinson, H. F. (1959). *Pleiotropism and the genetic variance covariance. Biometrics, 15,* 518–537.

Orf , J. H., Diers, B. W., & Boerma, H. R. (2004). Genetic Improvement: Conventional and Molecular-Based Strategies. In H. R. Boerma & J. E. Specht (3rd Ed.). *Soybeans: Improvement, Production, and Uses* (pp. 417-445). Madison, WI: American Society of Agronomy.

Probst A. H., Laviolette F. A., Athow, K. L., & Wilcox, J. R. (1971). *Registration of Protana Soybean. Crop Science, 11,* 312.

SAS Institute, Inc. Version [9.1.3] of the SAS System for [Windows]. Copyright © [2003] SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

Silvela, L. & Diez-Barra, R. (1985). *Recurrent Selection in Autogamous Species Under Forced Random Mating. Euphytica, 34,* 817-832.

Silvela, L., De La Pena, I., & Gomez-Ruano, R. (1999). *Selection under Negative Linkage Disequilbrium. Random Mating versus Inbreeding. Heredity, 82,* 598-604.

Smith, T. J., & Camper, H. M. (1973). *Registration of Essex Soybean (Reg. No. 97). Crop Science, 35,* 524-528.

Soliman, K. M., & Allard, R. W. (1991). *Grain Yield of Composite Cross Populations of Barley: Effects of Natural Selection. Crop Science, 31,* 705-708.

Stam P. (1977). *Selection Response Under Random Mating and Under Selfing in the Progeny of a Cross of Homozygous Parents. Euphytica, 26,* 169-184.

Stefaniak, T. R., Hyten, D. L., Pantalone, V. R., Klarer, A., & Pfeiffer, T. W. (2006). *Soybean Cultivars Resulted from More Recombination Events Than Unselected Lines in the Same Population. Crop Science, 46,* 43-51.

Suneson, C. A. (1956). *An Evolutionary Plant Breeding Method. Agronomy Journal, 48,* 188-191.

Tinius C. N., Burton, J. W., & Carter, T. E. Jr. (1991). *Recurrent Selection for Seed Size in Soybean: I. Response to Selection in Replicate Populations. Crop Science, 31,* 1137-1141.

Tinius C. N., Burton, J. W., & Carter, T. E. Jr. (1993) *Recurrent Selection for Seed Size in Soybean: III. Indirect Effects on Seed Composition. Crop Science, 33,* 959-962.

Usten, A., Allen, F. L., & English, B. C. (2001). *Genetic Progress in Soybean of the U.S. Midsouth. Crop Science, 41,* 993–998.

Wilcox, J.R. (2001). *Sixty Years of Improvement in Publicly Developed Elite Soybean Lines. Crop Science, 49,* 1711-1716.

Wilcox, J.R. (2004). World Distribution and Trade of Soybeans. In H.R Boerma and J. E. Specht (3rd Ed.). *Soybeans: Improvement, Production, and Uses* (pp. 1-14). Madison, WI: American Society of Agronomy.

Wilson, R.F. (2004). Seed Composition. In H.R Boerma and J. E. Specht (3rd Ed.). *Soybeans: Improvement, Production, and Uses* (pp. 621-678). Madison, WI: American Society of Agronomy.

Wolf, R. B., Cavins, J. F., Kleian, R., and Black, L. T. (1982). *Effect of Temperature on Soybean Seed Constituents: Oil, Protein, Mositure, Fatty Acids, Amino Acids, and Sugars. Journal of the American Chemists' Society, 59,* 230-232.

Table 1. Expected Mean Squares for the Analysis of Variance of each population for two replications grown at Plymouth, Clinton, and Mount Olive, NC in 2009.

Source of Variation	Expected Mean Square
Location	$\sigma^2 + r g \sigma^2_L + r g \sigma^2_L$
Replication(Location)	$\sigma^2 + g \sigma^2_{R(L)} + g \sigma^2_{R(L)}$
Genotype	$\sigma^2 + r \sigma^2_{LG} + r l \sigma^2_G$
Location*Genotype	$\sigma^2 + r \sigma^2_{LG}$
Error	σ^2

Number of Replications= r =2

Number of Locations= l =3

Number of Genotypes = g

Table 2. Means and Ranges with LSD for yield, flowering date, maturity date, lodging, height, seed weight, percent protein and percent oil, on a dry matter basis for two replications of 119 soybean lines for Population RSII grown at Clinton, Plymouth, and Mount Olive, NC in 2009.

Trait of Interest	Mean	High	Low	LSD
Yield in kg/ha	2070	2826	1415	620
Flowering Date from July 1	4.7	17.5	-3.8	1.7
Maturity Date from October 1	26.4	34.8	13.7	3.4
Lodging	2.8	4.6	1.1	0.7
Plant Height in cm	108	141	58	14
Seed Weight in g	14.3	16.9	11.7	1.6
Percent Protein	41.6	43.7	39.6	1.3
Percent Oil	19.3	20.7	17.9	1

Table 3. Means and Ranges with LSD for yield, flowering date, maturity date, lodging, height, seed weight, percent protein and percent oil, on a dry matter basis for two replications of 103 soybean lines for Population RSIII grown at Clinton, Plymouth, and Mount Olive, NC in 2009.

Trait of Interest	Mean	High	Low	LSD
Yield in kg/ha	2205	3142	1625	528
Flowering Date from July 1	5.6	19.5	-3.8	3.8
Maturity Date from October 1	26.8	35.5	13.5	3.1
Lodging	2.7	4.2	1.5	0.6
Plant Height in cm	111	131	88	11
Seed Weight in g	13.9	19.9	11.2	1.1
Percent Protein	41.6	45.1	39.3	1.3
Percent Oil	19	21	15.9	0.8

Table 4. Mean Squares, R^2, and CV for yield, flowering date, maturity date, lodging, height, seed weight, percent protein and percent oil for two replications of 119 soybean lines for Population RSII grown at Clinton, Plymouth, and Mount Olive, NC in 2009.

		Mean Squares							
Source	df	Yield (kg/ha)	Flowering Date (days)	Maturity Date (days)	Lodging (1 to5)	Height (cm)	Seed Weight (grams)	Protein (%)	Oil (%)
Loc	2	38234134	7824.7	949.4	33	13851	110.2	391	52.1
Rep(loc)	3	1668779	10.2	0.25	0.6	630	1.7	22.9	8.6
Geno	117	4362559**	34.2**	71.3**	2.6**	869**	7.1**	4.1**	1.9**
Loc*geno	223	299913**	2.2**	8.9**	0.4**	158**	1.9**	1.4**	0.8**
Error	318	195194	4.9	4.6	0.3	93	0.9	0.5	0.4
R^2		0.8	0.9	0.9	0.9	0.85	0.84	0.91	0.81
CV		21.7	8.9	8.2	18.9	8.9	6.6	1.71	3.28

*, significance at $\alpha = 0.05$; **, significance at $\alpha = 0.01$.

Table 5. Mean Squares, R^2, and CV for yield, flowering date, maturity date, lodging, height, seed weight, percent protein and percent oil for two replications of 103 soybean lines for Population RSIII grown at Clinton, Plymouth, and Mount Olive, NC in 2009.

		Mean Squares							
Source	df	Yield (kg/ha)	Flowering Date (days)	Maturity Date (days)	Lodging (1 to5)	Height (cm)	Seed Weight (grams)	Protein (%)	Oil (%)
Loc	2	16812683	7368.3	902	31.6	20974	171.8	207.5	10.4
Rep(loc)	3	4636665	13.9	38.1	4.7	422	22.2	5.0	3.3
Geno	102	480692**	174**	80.3**	1.4**	498**	18**	9.8**	5**
Loc*geno	204	217968**	11.5**	7.7**	0.3**	91**	0.9**	1.3**	0.5**
Error	297	196325	4.6	5.3	0.3	73	0.5	0.7	0.3
R^2		0.71	0.96	0.88	0.79	0.84	0.95	0.89	0.88
CV		20.09	8.36	8.62	18.75	7.72	4.98	1.99	2.81

*, significance at $\alpha = 0.05$; **, significance at $\alpha = 0.01$.

Table 6. Genotypic Variance, GxE Variance, Phenotypic Variance, Heritability Estimates with Confidence Intervals for yield, flowering date, maturity date, lodging, height, seed weight, percent protein and percent oil of 119 soybean lines for Population RSII grown at Clinton, Plymouth, and Mount Olive, NC in 2009 with two replications.

Trait of Interest	G.C.V.	Geno Var (SE)	GxE Var (SE)	Pheno Var (SE)	Heritability w/C.I.
Yield (kg/ha)	7.7	25351.4 (10798.7)	51308.5 (7350.1)	76598.1 (9716.2)	0.331 (0.48, 0.104)
Flowering Date	112	27.4 (0.762)	2.8 (0.09)	29.1 (0.762)	0.941 (0.952, 0.918)
Maturity Date	12.8	11.4 (1.59)	2.3 (0.213)	12.9 (1.59)	0.882 (0.905, 0.836)
Lodging	22.5	0.4 (0.058)	0.1 (0.01)	0.5 (0.057)	0.849 (0.877, 0.788)
Plant Height (cm)	10.4	126.4 (19.52)	30.9 (3.82)	152.8 (19.356)	0.827 (0.862, 0.762)
Seed Weight (g)	6.7	0.9 (0.161)	0.5 (0.045)	1.2 (0.16)	0.734 (0.797, 0.651)
Percent Protein	1.7	0.5 (0.094)	0.5 (0.02)	0.7 (0.092)	0.675 (0.751, 0.57)
Percent Oil	2.2	0.2 (0.045)	0.2 (0.02)	0.3 (0.043)	0.56 (0.684, 0.456)

Table 7. Genotypic Variance, GxE Variance, Phenotypic Variance, Heritability Estimates and Confidence Intervals for yield, flowering date, maturity date, lodging, height, seed weight, percent protein and percent oil of 103 soybean lines for Population RSIII grown at Clinton, Plymouth, and Mount Olive, NC in 2009 with two replications.

Trait of Interest	G.C.V.	Geno Var (SE)	GxE Var (SE)	Pheno Var (SE)	Heritability w/ C. I.
Yield (kg/ha)	9.6	44976.6 (11672)	11719.4 (5981)	81740.9 (11110)	0.55 (0.661, 0.403)
Flowering Date	95.2	28.2 (4.03)	3.4 (0.276)	30.1 (4.02)	0.938 (0.951, 0.913)
Maturity Date	13.1	12.4 (1.86)	1.1 (0.198)	13.6 (1.856)	0.907 (0.928, 0.874)
Lodging	15.8	0.2 (0.032)	0.02 (0.008)	0.2 (0.031)	0.781 (0.841, 0.72)
Plant Height (cm)	7.5	68.5 (11.59)	9.7 (2.416)	83.8 (11.5)	0.817 (0.864, 0.76)
Seed Weight (g)	12.3	2.9 (0.417)	0.2 (0.022)	3.1 (0.416)	0.949 (0.964, 0.936)
Percent Protein	2.9	1.5 (0.228)	0.3 (0.031)	1.7 (0.227)	0.874 (0.905, 0.833)
Percent Oil	4.7	0.8 (0.118)	0.1 (0.011)	0.9 (0.117)	0.913 (0.933, 0.882)

Table 8. Rank correlations of two replications of 119 soybean lines in Population RSII and 103 soybean lines in Population RSIII by location for Clinton, Plymouth, and Mount Olive, NC in 2009.

RSII				RSIII		
	Clinton	Mt. Olive	Plymouth	Clinton	Mt. Olive	Plymouth
Clinton		0.0401	0.319**		0.0817	0.2295*
Mt. Olive			0.1202			0.4229**

*, significance at $\alpha = 0.05$; **, significance at $\alpha = 0.01$.

Table 9. Yield Mean Rankings of those genotypes with overlap in the top and bottom 20% of two replications of 119 soybean lines in Population RSII and 103 soybean lines in Population RSIII by location for Clinton, Plymouth, and Mount Olive, NC in 2009.

RSII				RSIII			
Genotype	Clinton	Mt. Olive	Plymouth	Genotype	Clinton	Mt. Olive	Plymouth
N-202	18	107	105	N-353	74	102	19
N-204	7	106	9	N-376	83	21	59
N-222	87	13	101	N-379	21	40	90
N-236	117	4	61	N-380	93	12	24
N-238	106	9	38	N-385	90	77	3
N-265	25	105	7	N-387	99	16	45
N-266	99	93	24	N-398	15	94	75
N-284	13	108	22	N-404	7	84	64
N-299	109	1	53	N-421	10	81	83
N-319	113	3	95	N-426	91	5	13
N-331	97	96	1	N-431	84	9	68
				N-454	103	8	103
				N-493	17	30	82

Table 10. Genotypic Correlations with their Standard Errors and Phenotypic Correlations (in parenthesis) between all 28 pairs of traits for two replications of 119 soybean lines from Population RSII grown at Clinton, Plymouth, and Mount Olive, NC in 2009.

	Flowering	Maturity	Lodging	Seed Weight	Protein	Oil	Height	Yield
Flowering		0.45 ± 0.02 (0.46±0.02)	0.3 ± 0.03 (0.3±0.02)	0.11 ± 0.05 (0.1±0.04)	-0.15 ± 0.05 (-0.14±0.04)	0.07 ± 0.06 (0.06±0.04)	0.42 ± 0.03 (0.42±0.02)	-0.23 ± 0.08 (-0.13±0.04)
Maturity			-0.1 ± 0.17 (-0.07±0.14)	0.43 ± 0.09 (0.41±0.07)	-0.15 ± 0.11 (-0.11±0.09)	0.08 ± 0.13 (0.06±0.09)	0.23 ± 0.1 (0.21±0.08)	0.29 ± 0.16 (0.23±0.09)
Lodging				-0.04 ± 0.11 (-0.02±0.09)	-0.04 ± 0.12 (-0.02±0.09)	0.24 ± 0.13 (0.17±0.09)	0.43 ± 0.08 (0.41±0.06)	-0.39 ± 0.17 (-0.16±0.09)
Seed Weight					0.01 ± 0.13 (0.07±0.09)	0.18 ± 0.14 (0.12±0.09)	0.22 ± 0.11 (0.18±0.09)	0.24 ± 0.18 (0.24±0.09)
Protein						-0.61 ± 0.09 (-0.64±0.06)	-0.02 ± 0.12 (-0.00±0.09)	-0.17 ± 0.2 (0.05±0.09)
Oil							0.18 ± 0.13 (0.12±0.09)	0.07 ± 0.23 (-0.01±0.09)
Height								0.16 ± 0.17 (0.14±0.09)
Yield								

Significant

Table 11. Genotypic Correlations and Phenotypic Correlations (in parenthesis) between all 28 pairs of traits for two replications of 103 soybean lines from Population RSIII grown at Clinton, Plymouth, and Mount Olive, NC in 2009.

	Flowering	Maturity	Lodging	Seed Weight	Protein	Oil	Height	Yield
Flowering		0.43 ± 0.09 (0.4±0.08)	0.35 ± 0.08 (0.33±0.07)	-0.17 ± 0.1 (-0.17±0.1)	-0.22 ± 0.1 (-0.21±0.09)	-0.01 ± 0.11 (-0.0±0.1)	0.45 ± 0.09 (0.39±0.08)	-0.43 ± 0.12 (-0.3±0.09)
Maturity			-0.05 ± 0.12 (-0.02±0.1)	0.28 ± 0.1 (0.27±0.09)	-0.21 ± 0.11 (-0.19±0.09)	0.14 ± 0.11 (0.13±0.1)	0.17 ± 0.11 (0.17±0.1)	0.19 ± 0.14 (0.14±0.1)
Lodging				-0.27 ± 0.11 (-0.23±0.09)	-0.2 ± 0.12 (-0.14±0.1)	0.01 ± 0.12 (0.01±0.1)	0.29 ± 0.12 (0.27±0.09)	-0.6 ± 0.13 (-0.37±0.08)
Seed Weight					0.02 ± 0.11 (0.09±0.16)	0.18 ± 0.1 (0.15±0.1)	0.14 ± 0.11 (0.13±0.1)	0.63 ± 0.1 (0.48±0.07)
Protein						-0.73 ± 0.05 (-0.72±0.04)	-0.08 ± 0.12 (-0.05±0.1)	0.07 ± 0.14 (0.09±0.1)
Oil							-0.04 ± 0.12 (-0.05±0.1)	0.26 ± 0.14 (0.15±0.1)
Height								-0.21 ± 0.15 (-0.08±0.1)
Yield								

Significant

Table 12. Gene frequencies and Chi-squared test of the parents and progeny of 119 soybean lines in Population RSII and 103 soybean lines in Population RSIII by location for Clinton, Plymouth, and Mount Olive, NC in 2009.

	RSII				RSIII			
	White	Purple	Tawny	Gray	White	Purple	Tawny	Gray
Parents	0.58	0.42	0.40	0.60	0.75	0.25	0.33	0.67
Progeny	0.53	0.47	0.80	0.20	0.51	0.49	0.58	0.42
	Flower Color		Pubescence Color		Flower Color		Pubescence Color	
Chi-squared	1.03		66.67**		30.72**		28.27**	

**, significance at α= 0.01.

Table 13. Direct and Indirect Response to Selection for yield, protein, and oil of 119 soybean lines in Population RSII and 103 soybean lines in Population RSIII by location for Clinton, Plymouth, and Mount Olive, NC in 2009.

	RII				RSIII			
	Direct	Indirect			Direct	Indirect		
		Yield	Percent Oil	Percent Protein		Yield	Percent Oil	Percent Protein
Yield (kg/ha)	160.57	-	0.03	-0.122	272.15	-	0.292	0.105
Percent Oil	0.565	14.417	-	-0.558	1.47	89.79	-	-1.456
Percent Protein	1.008	-39.508	-0.377	-	1.956	23.17	-1.048	-

APPENDICES

APPENDIX A

Additional Figures

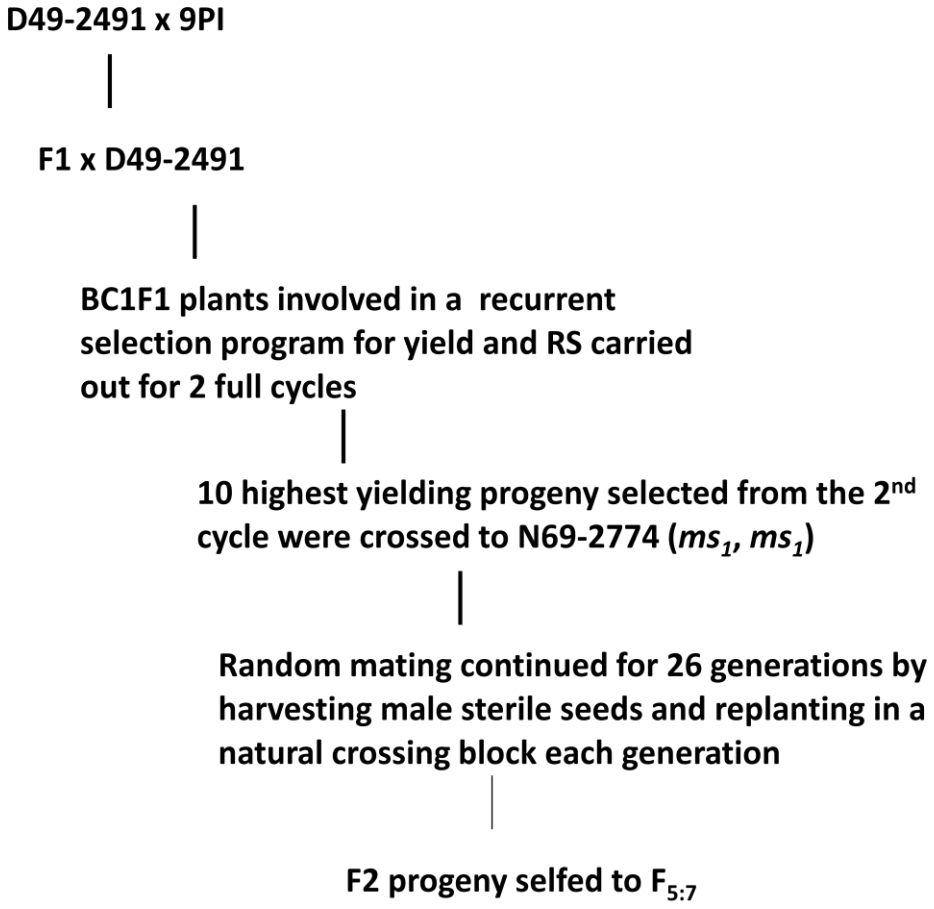

Figure A1. Population Development of RSII

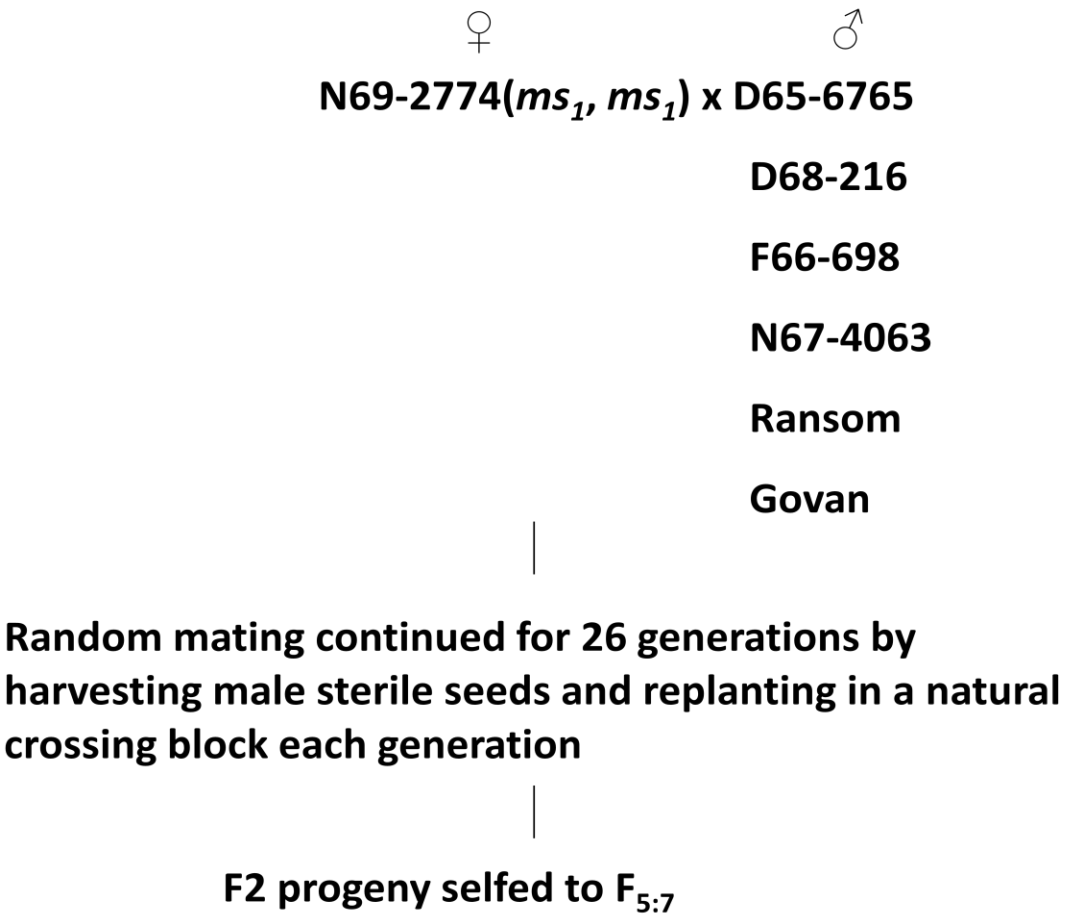

Figure A2. Population Development of RSIII

APPENDIX B

Additional Tables

Table B1. Parent means for yield, flowering date, maturity date, lodging, height, seed weight, percent protein and percent oil from Population RSII, taken from the GRIN database.

	Trait of Interest								
Parents	Maturity	Lodging	Seed Weight	Protein	Oil	Height	Yield	Flower	Pubescence
PI 96035	VI	2	11.5	47.7	16.1	68	1220	W	G
PI 171445	VII	4	12.3	47.2	15.7	100	1900	P	T
FC 31649	VII	1.5	16.1	45.3	19.9	97	1870	W	G
PI 165675	VII	4	11.7	45.2	19.3	100	2210	P	T
PI 165673	VI	3	10.7	51.5	11.7	106	760	P	G
FC 31676	VII	1	15.6	45.9	20.3	91	1750	W	G
PI 153681	VII	3	14.1	43.3	20.4	133	1390	W	G
PI 97100	VII	3	11.7	45.9	18.5	135	1510	W	G
PI 54618-1	VI	unknown	18.0	39.4	21.9	unknown	unknown	W	G
D49-2491	VI	1.9	12.8	40	21.8	77	2047	P	T
N69-2774	unknown	unknown	unknown	unknown	unknown	unknown	unknown	W	G

Table B2. Parent means for yield, flowering date, maturity date, lodging, height, seed weight, percent protein and percent oil from Population RSIII, taken from previous studies.

	Trait of Interest								
Parents	Maturity	Lodging	Seed Weight	Protein	Oil	Height	Yield	Flower	Pubescence
Ransom	VII	1.5	15.6	39.6	23.7	81	2441	P	T
Govan	VII	1	11.3	45.5	19.5	99	2450	W	G
D65-6765	VII	1.6	12.1	44.2	19.5	86	2024	W	T
F66-698	VII	2.3	14.9	44.3	19.9	91	2668	P	T
D68-216	VI	2.5	unknown	40.5	20.6	86	2204	W	T
N67-4063	VI	unknown	unknown	40.8	22.3	91	2426	P	G
N69-2774	unknown	unknown	unknown	unknown	unknown	unknown	unknown	W	G

Table B3. Yield Means for two replications of 119 soybean lines for Population RSII grown at Clinton, Plymouth, and Mount Olive, NC in 2009.

Genotype	Yield
N-197	1965.64
N-198	2184.37
N-199	2603.33
N-200	2214.58
N-201	1967.46
N-202	1860.43
N-204	2215.67
N-206	1985.32
N-207	1622.23
N-210	2071.72
N-211	1985.45
N-214	1759.38
N-215	1959.35
N-217	2214.34
N-220	2717.34
N-222	1893.74
N-223	2225.3
N-224	1902.45
N-226	2023.91
N-229	1782.36
N-232	2530.25
N-233	1738.34
N-234	2484.57
N-235	2123.16
N-236	2004.29
N-237	2393.76
N-238	2096.86
N-239	2825.74
N-240	1776.59
N-241	2427.27
N-242	1798.12
N-246	2047.93
N-247	2113.23
N-249	1909.67
N-250	1941.13
N-252	2214.06
N-253	2098.98

Table B3 Continued

N-255	1672.5
N-256	2538.25
N-257	2245.22
N-258	1737.2
N-259	2307.4
N-260	1765.03
N-261	1787.51
N-262	2149.58
N-263	1621.19
N-264	2225.39
N-265	2152.56
N-266	1856.97
N-267	2339.58
N-268	2018.8
N-269	1824.9
N-270	2303.67
N-271	1859.97
N-272	2607.21
N-273	1989.78
N-274	1881.51
N-275	1438.27
N-276	2179.16
N-278	2156.81
N-280	1919.93
N-281	2248.46
N-283	2040.83
N-284	2193.2
N-285	2205.15
N-286	1655.55
N-288	1632.69
N-289	2238.92
N-290	2248.82
N-291	2023.04
N-292	2566.85
N-293	1999.74
N-294	1919.79
N-295	1852.23
N-296	2079.82
N-297	2012.54
N-299	2041.32
N-300	1811.92
N-301	2312.25
N-302	1989.78
N-303	2183.42

Table B3 Continued

N-304	1753.15
N-305	1914.76
N-306	1602.37
N-307	2241.37
N-308	2174.15
N-309	1532.48
N-310	1771.04
N-311	1814.83
N-313	1876.66
N-316	2432
N-318	1865.34
N-319	1808.49
N-320	2588.72
N-321	2386.34
N-322	2184.38
N-323	1796.86
N-324	2008.16
N-325	2259.48
N-326	1992.91
N-327	1952.59
N-328	2456.31
N-329	1530.98
N-330	2252.63
N-331	2090.97
N-333	2119.7
N-334	1414.89
N-335	2684.03
N-336	2499.57
N-337	2350.03
N-338	2320.75
N-339	2397.49
N-342	2061.65
N-343	1944.27
N-344	1871.66
N-346	2578.07
N-348	2111.16
N-349	2149.63
Raleigh	2745.38

Table B4. Yield Means for two replications of 103 soybean lines for Population RSIII grown at Clinton, Plymouth, and Mount Olive, NC in 2009.

Genotype	Yield
N-350	2770.03
N-351	1995.18
N-352	2504.05
N-353	1960.7
N-354	1893.52
N-355	2520.02
N-356	2597.67
N-358	2692.66
N-359	2035.36
N-361	2498.32
N-363	2438.5
N-364	2300.91
N-365	2234.25
N-366	2223.66
N-367	1998.82
N-368	2085.44
N-372	1830.42
N-373	2226.39
N-374	1919.45
N-375	2033.8
N-376	2218.14
N-377	1909.7
N-378	2211.7
N-379	2217.77
N-380	2311.14
N-381	1894.96
N-383	2457.68
N-384	2167.2
N-385	2162.92
N-386	2373.16
N-387	2173.03
N-388	2005.33
N-389	2557.47
N-390	2238.96
N-391	2005.9
N-392	1674.44

Table B4 Continued

N-394	2207.64
N-396	2306.34
N-397	2358.11
N-398	2122.23
N-399	2602.05
N-400	1840.25
N-401	2360.27
N-402	2320.53
N-403	1730.43
N-404	2265.93
N-405	1625.28
N-406	1921.09
N-408	2780.77
N-409	1958.67
N-411	2329.14
N-412	3142.34
N-413	2044.09
N-414	2170
N-416	2004.77
N-417	2039.91
N-419	1973.1
N-420	2460.82
N-421	2031.45
N-423	2646.52
N-426	2462.28
N-427	2346.77
N-428	2099.87
N-429	2428,02
N-430	1865.29
N-431	2120.62
N-432	2067.27
N-433	2164.91
N-434	1706.04
N-435	2174.1
N-437	1965.48
N-438	1918.12
N-441	1682.2
N-442	1809.87
N-443	2375.57
N-444	2702.24

Table B4 Continued

N-445	2283.89
N-447	2383.89
N-449	2412.64
N-450	2009.98
N-451	2587.85
N-453	2421.03
N-454	1695.37
N-456	2229.03
N-460	2446.32
N-461	2311.28
N-462	2209.54
N-463	2626.53
N-464	2062.72
N-465	1763.01
N-469	2315.22
N-471	2145.52
N-473	2166.97
N-476	2058.71
N-477	2860.73
N-480	2274.95
N-481	2004.71
N-482	1963.82
N-483	2045.22
N-487	1932.43
N-491	2460.87
N-493	2340.88
N-494	2651.98
Roy	3306.15
Raleigh	3003.98

APPENDIX C

Linkage Phase Disequilibrium

In these two populations, random mating had occurred for 26 generations. Taking the theoretical initial LD values of 0.25, 0.1, and 0.01 (0.25 being the maximum value) and recombination values at increments of 0.005 starting at 0.5 and going to 0.005, current LD values were calculated for each recombination rate based on the 26 generation of random mating. Any current LD value of 0.01 was considered to have reached linkage equilibrium. The following equation was used to determine current LD:

$D_n=(1-r)^nD_0$

where D_n is the current theoretical value of D after n generations of random mating, D_0 is the initial value of D and r is the recombination rate.

TableC1. D_n values after 26 generations of random mating with a theoretical D_0 value.

Recombination Rate for equilibrium (r)	**Theoretical D_0 values**		
	D_0=0.25	D_0=0.1	D_0=0.01
0.105	0.0139	0.005	0.001
0.075	0.0329	0.0132	0.0013
0.005	0.219	0.0877	0.009

For the theoretical maximum D value of 0.25, linkage equilibrium was achieved for loci with recombination rates greater than 0.105. Any loci under tight linkage had not achieved linkage equilibrium in these populations. For the theoretical D value of 0.1, linkage equilibrium had been achieved for any loci with a recombination rate great than 0.075. For this value of D,

extremely tight linkages, those with a distance of 7 cM, will still have some LD. Under the theoretical D value of 0.01, which is very small and near equilibrium, any loci with a recombination rate greater than 0.005 were considered to be under linkage equilibrium. Therefore all loci would be in linkage equilibrium under this theoretical LD value. In these populations the most probable initial LD value would be between 0.25 and 0.1 based on the fact that soybeans are naturally self pollinated and would likely exhibit higher D values. So in these populations, it is likely that loci with recombination rates greater than 0.075 could be in linkage equilibrium. Leaving only the very tight gene linkages in linkage disequilibrium in these populations.

With the theoretical initial D value of 0.25, and 100 generations of random mating, linkage equilibrium will be achieved for those loci which have a recombination rate greater than 0.03, or 3 cM distance. After only 50 generations loci with a recombination rate greater that 0.055 will be in linkage equilibrium.