ABSTRACT

FRIEDRICHS, MARTIN RAY. Heterosis and Inbreeding depression in *Glycine max*. (Under the direction of Dr. Joseph Burton.)

This study focused on soybean heterosis in regard to yield, seed size, and height and consisted of three objectives: (1) to study three new hybrids and retest a fourth hybrid for heterosis, (2) to study the correlation between heterosis in the F_2 generation and inbred line variance, (3) to find SSR markers that help to explain heterosis in a heterotic hybrid.

Objective one was evaluated in 2007 and 2008. The four hybrids were tested in yield trials that consisted of F_1 - F_4 generations and parents. Results showed some heterosis, but there was a lack of consistency across years due to a genotype by year effect.

The second objective was evaluated in 2008. Yield trials examining six populations were tested over three locations. The hybrid seed were developed in 2007 and F_1 plants were grown to generate F_2 bulks, the inbred lines were developed by single seed descent and tested as $F_{6:8}$ lines. A correlation between heterosis and transegregate lines was found for yield and seed size.

For the third objective to identify SSR markers for heterosis, the cross of Hutcheson x Holladay was used to develop $F_{2:3}$ and $F_{2:4}$ lines. A DNA sample was taken from 360 individual F_2 plants and screened with 40 polymorphic markers that represented each arm of the 20 linkage groups. Yield, seed size and height were taken on the $F_{2:3}$ and $F_{2:4}$ lines and used to evaluate the marker data. No consistent marker for additive and dominance variance was found between the generations.

Heterosis and Inbreeding Depression in *Glycine max*

by
Martin Ray Friedrichs

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Crop Science

Raleigh, North Carolina

2009

APPROVED BY:

_____	_____
Dr. Thomas E. Carter, Jr.	Dr. Joseph Burton Committee Chair
_____	_____
Dr. David Dickey	Dr. James Holland

BIOGRAPHY

Martin R. Friedrichs was born in Pittsburg, Kansas on March 6th, 1981. Growing up on a dairy farm, he developed an interest in horticulture. He graduated from Groton High School, Groton, South Dakota, in 2000. He attended Southwest Missouri State University, where he received his B.S. degree in Horticulture in 2004. While attending SMSU, he met and married his wife, Aja.

During his undergraduate studies, Martin worked as an intern at the United States Department of Agriculture Agricultural Research Service in Raleigh, North Carolina in the summer and fall of 2003, developing his interest in plant breeding. In the fall of 2005, Martin was accepted as a graduate student at North Carolina State University under the guidance of Dr. Joseph Burton. In March of 2009, Martin began working as a Sweet Corn breeder at the Syngenta research station in Naples, Florida.

ACKNOWLEDGEMENTS

I wish to express my gratitude to my major professor, Dr. Joseph W. Burton, for his support and direction during the project. His patience, sincerity, and support throughout my time as an intern and graduate student are greatly appreciated. It has been a great pleasure working with him.

I would also like to thank my committee members: Dr. Thomas Carter, Dr. David Dickey, and Dr. James Holland, for their assistance and insights these research projects. It has been a privilege collaborating with each of you. I would also like to thank Dr. Andrea Cardinal for allowing the use of her laboratory. I am also grateful to her staff, which provided valuable technical assistance.

In addition, I would like to thank the rest of the USDA soybean breeding staff: William Novitzky, Dr. Jesse Gilsinger, Earl Huie, Dr. Margarita Villagarcia, Bobby McMillen, Fred Farmer, John Graber, Phil Rzewnicki, Austin Hodges, and the part-time employees. Your help throughout these projects is greatly appreciated.

Finally, my thanks to my wife, Aja, for allowing me to pursue my dream. Without her support and encouragement, I could not have succeeded.

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES

INTRODUCTION

Soybean Importance

Nearly all soybeans are processed for their oil. This oil can then be refined for cooking or other edible uses, or utilized in various industrial applications, including biodiesel. The high protein meal is processed primarily for use in animal feed, but it is also used in various food products. New, innovative uses for soy constituents are being developed such as the use of soybean fiber for plywood and soy oil for inks and crayons.

Soybeans, in the United States were planted on 63.6 million acres (25.7 million hectares) in 2007, producing 2.585 billion bushels (70.36 million metric tons) of soybeans. Total world production of soybeans for 2007 was 8.07 billion bushels. In 2007, soybeans represented 56 percent (219.9 million metric tons) of world oilseed production, the next largest oil producers were rapeseed and cottonseed at 12 percent (Soystat.com).

Soybean Breeding Goals

In addition to increasing the crop productivity, maintaining or increasing protein and oil concentration are ongoing breeding goals. While an increase in both is ideal, protein and oil are negatively correlated; thus, an increase in one usually results in a decrease of the other (Brim and Burton, 1979). Additionally, the profile of the fatty acids is important. Reducing linolenic acid and palmitic acid are also breeding goals. The reduction of palmitic acid will make the oil healthier while a reduction in linolenic

increases the stability of the oil in frying applications (Liu and White, 1992). The quality of the protein also is being addressed. Soy protein is low in sulfur amino acids, methionine and cystine, which must be added to the soy meal to meet dietary needs of livestock.

Soybean Breeding Methods

The prevailing method of soybean breeding is a biparental cross, followed by inbreeding to F_3, F_4 or F_5, before deriving lines. The inbred lines are then tested in multiple environments for one or more years. Inbreeding is achieved through single seed or pod descent. A single seed (for practical reasons a single pod) is picked per plant with one seed planted for the next generation. The second seed in each pod is saved in case the first planting fails. This process is continued until the desired level of inbreeding is reached at which point each progeny which traces back to a single F_2 plant is maintained in bulk. Soybean breeders have used a multiple-seed procedure in order to avoid high labor costs. Instead of taking one seed per plant, one pod with two or three seed is picked and bulked within segregating populations.

In addition to the phenotypic selection done upon the single seed descent derived inbred lines, marker technology is being utilized to increase the precision and speed of selection. Major breeding companies now use disease resistance and Genetically Modified Organism (GMO) markers to select the most desirable lines during inbreeding.

Soybean Reproduction

As a result of the propensity of *Glycine max* to self pollinate, all commercially available lines are inbred lines. Nearly all soybean breeding is conducted by creating and testing inbred lines. Because of this emphasis on inbred lines, hybrid breeding has been neglected.

Heterosis Hypothesis

Heterosis (or hybrid vigor) is the superiority the F_1 generation in relation to their parents (Fehr, 1987). First coined by G.H. Shull in 1914, heterosis has been exploited by breeders to enhance the productivity of numerous crop and horticultural plants. The effects of the phenomenon have been quantified in a wide variety of plant studies (Stuber, 1994).

There are two different measures of heterosis. Mid-parent heterosis is defined as the increased vigor of the F_1 over the mean of the parents. High-parent heterosis is defined as the increased vigor of the F_1 over the greater parent (Crow 1999). We will be mainly concerned with high-parent heterosis because mid-parent heterosis will always be smaller or equivalent to high-parent heterosis if parents are equal (Crow 1999). There are three historically important theories explaining heterosis: dominance, overdominance, and

epistasis.

The dominance hypothesis assumes that, in hybrids, detrimental recessive alleles from one parent are masked by dominant alleles from the other. The overdominance hypothesis assumes that, for at least some loci, heterozygotes are superior to either homozygote. The epistasis hypothesis proposes that the combinations of additive and dominant alleles result in an increase in hybrid performance. These three hypotheses have been treated as oppositional, but they are not mutually exclusive. In fact they are all consistent with heterosis and inbreeding depression phenomena; distinguishing between them can be difficult (Crow 2000).

From the time of Jones (1918) and Collins (1921) until the late 1940's, dominance was the accepted explanation for heterosis. In 1945, Hull made three arguments for overdominance in maize that brought it back as a potential explanation for heterosis. First, he argued that the failure of mass selection to produce substantial increases in yield pointed to the failure of the dominance hypothesis. Second, he argued that the yield of most hybrids was greater than the sum of the inbred parents and this could not be explained through simple dominance. However, his third argument carried the most weight. He noted what he called the "Fisher Equilibrium" (Fisher 1922), the maintenance of stable polymorphisms under overdominance. His argument utilized what he called "constant parent regression." The regression of the F_1 on the value of a parent had different expectations of dominance and overdominance, when the other parent was held constant. When constant parent was high-yielding, the regression could curve and be

negative for some values, which is consistent with overdominance (Crow 2000).

Comstock and Robinson (1952) devised three breeding experiments for measuring average additive and dominance effects of genes. From these experiments, average dominance could be estimated and compared to additive effects. In experimental design III, a random sample of F_2 individuals derived from a cross between two homozygous inbred lines is backcrossed to each of the parents, yielding a population with gene and genotype frequencies equivalent to an F_2. A one-way analysis of variance (ANOVA) of phenotypic means and differences of the F_2 backcrosses yields estimates of the dominance and additive genetic variance with nearly equal precision and their ratio provides a weighted estimate of the squared degree of dominance. In Experiment III the reported dominance value was 1.6, which would point to overdominance. A measure of 0 means no dominance, a measure of 1 means complete dominance and a measure above 1 means overdominance. When Experiment III is continued for successive cycles, recombination occurs and allows linkage disequilibrium to dissipate and decrease the repulsion-phase linkages (Crow 2000). Gardner (1963) tested four cycles of Experiment III. In the first cycle, the dominance value, 1.4, was in the overdominance range. In the second generation, the dominance value was just above 1. By the third and fourth generation, these dominance values had fallen well into the partial dominance range (any value between 0 and 1). The results of this experiment showed that the apparent overdominance of the early generations was in fact pseudo- overdominance due to linkage disequilibrium.

Epistasis is the final possibility for heterosis. Unique epistatic combinations of

alleles formed as a result of a cross could show heterotic increases in yield. There has, however, been justification for the convenient custom of ignoring epistasis. Kimura (1965) showed that if the recombination value is larger than the selective difference (which is true for most pairs of loci) the population attains an approximately constant level of linkage disequilibrium after several generations of selection; he called this *quasi-linkage equilibrium*. In this state, the variance due to linkage disequilibrium and that due to epistasis cancel almost exactly. Thus progress under selection is given by the additive variance alone, with no appreciable contribution form epistatic components (Kimura 1965) and therefore justification for ignoring epistasis.

Molecular Genetics of Heterosis

Swanson-Wagner et al. (2006) used microarray technology to characterize the modes of gene action for 13,999 cDNAs in corn. Nearly 10% of the Expressed Sequence Tags (ESTs) exhibited differential expression among the three genotypes. Of the differentially expressed ESTs, the majority, 78%, were statistically indistinguishable from additivity. Most of the remaining 22% exhibited high parent dominance. However 2% exhibited low-parent dominance <1% exhibited underdominance, and 3% exhibited overdominance. These results with additive, dominant, and overdominant ESTs, is consistent with the hypothesis that multiple molecular mechanisms contribute to heterosis (Swanson-Wagner et al. 2006) However this experiment was done using F_1s so the result are not representative because F_2's are needed in order to have a segregating generation that

can be used to measure additive, dominance and overdominance effects.

Aside from the quantitative theories of heterosis (dominance, overdominance, and epistasis), physiological and biochemical explanations have been suggested. Milborrow (1998) postulated that hybrid vigor is a phenomenon in which strict regulatory limitation of growth is relaxed by heterozygosity. Variation for the presence of genes, the presence of novel beneficial alleles, and modified levels of gene expression in hybrids may all contribute to the heterotic phenotypes (Springer and Stupar, 2007). If growth is limited by the action of a number of randomly segregating regulatory factors, then recombining different homozygous strains in all combinations should occasionally bring together controlling factors which exert a stronger restrictive influence (subtractive heterosis) when present in a hybrid strain then when they are separate in their two homozygous parents. This 'subtractive heterosis' could be expected in a few F_1 hybrids (Milborrow 1998). Gowen (1952) shows an example where an F_1 hybrid exhibits subtractive heterosis in comparison with the mean of the two parents (Gowen 1952). Crow (1998) suggested that this type of data has been missed in the past due to the emphasis placed upon the positive side of heterosis. Any data showing 'subtractive heterosis' would have been discounted as having been caused by a disease or environmental conditions (Crow 1998).

With the advent of molecular tools, research has been conducted to analyze the different molecular aspects of heterosis. Colinearity has been studied as a possible explanation for heterosis. Therefore, gene deletions in inbred maize lines might have only minor quantitative effects on plant performance because these genes might often be

functionally compensated by duplicate copies elsewhere in the genome (Fu and Donner, 2002). The hemizygous complementation of many genes with minor quantitative effects in hybrids might lead to a significantly increased performance of hybrid plants and would be consistent with the dominance hypothesis. In soybeans, Schlueter et al., (2004) found that when using EST databases at least two genome duplication events occurred (Schlueter et al., 2004). Several previous studies had shown similar results with hybridization-based genetic maps (Shoemaker et al. 1996; Lee et al. 1999; Lee et al. 2001) and Restriction fragment length polymorphism (RFLP) probes which detected extensive homoeologous relationships between linkage groups (Shoemaker et al. 1996). This genome duplication would contribute to the conservation of mildly deleterious genes (Husband and Schemske, 1996) which would partly explain heterosis in an F_1 hybrid that combined favorable alleles at both loci, one contributed from each parent.

In maize, it is postulated that there may be heterosis-associated gene expression. This notion is supported by the observation that different tissues and organs within a hybrid plant display significant differences in their degree of heterosis (Melchinger 1999). However no consensus gene set was found that was differentially expressed between all inbred/hybrid combinations when analyzing more than one inbred versus hybrid combination (Uzarowska 2007, Guo 2006).

Allele-specific analyses can help to distinguish between cis- and trans-acting regulation of gene expression. A gene that is completely subjected to trans-regulation is expected to provide a similar contribution of both alleles to gene expression in the hybrid,

whereas genes subjected to cis-regualtion will show unequal expression of the two alleles in the hybrid (Wittkopp 2004). In a study done by Stupar and Springer (2006), 32 genes were analyzed. For 18 genes cis-regulation was prevailing, whereas only one gene was classified as trans-regulated. The remaining 13 genes were classified as a combination of cis- and trans-regulation. Differential response of the two alleles in a hybrid to environmental stress implied unequivalent functions of the different alleles that might have an impact on heterosis.

Experimental Evidence for Heterosis in Soybean

With current soybean breeding practices, heterosis is not being exploited in soybean. Typically, crosses are made, inbred lines are developed, and then yield testing is done using those inbred lines. Burton and Brownie (2006) speculated that the existence of heterosis in a particular cross combination may be evidence that superior gene combinations are possible in inbred lines derived from that cross.

Burton (2008) reviewed 9 experiments in which 260 F_1's from biparental crosses were compared with their parents in a spaced plant field design, the mean percentage high-parent heterosis of F_1 was 13.4% (Table 1). Fifty-five percent of those combinations had F_1 yields greater than the high-parent. The average percentage mid-parent heterosis of the F_1's was 25.7%. Cerna et al. (1997) tested 16 F_1's in a replicated spaced plant design (2 years, 2 locations) and found significant mid-parent and high-parent heterosis for 11 and 5 of the crosses, respectively.

Burton (2008) also reviewed several experiments conducted in standard experimental yield plots. An average of 20% high-parent heterosis for the F_1's of 2 crosses was found by Brim and Cockerham (1961); an average 6.2% high-parent heterosis among the F_1's of 8 crosses was reported by Hillsman and Carter (1981). Nelson and Bernard (1984) reported an average 3.3% high-parent heterosis among 37 crosses. Over these 3 experiments, 46 crosses were tested and 32 exhibited high-parent heterosis (Table 1.1). More recently Lewers et al. (1998) investigated 36 test crosses. They used 3 isolines of Clark and Harosoy as male parents and crossed these with 6 genetically diverse soybean lines as their female parents. The 36 test crosses were yield tested and an average F_1 mid-parent heterosis between 9.3% and 2.5% was reported. Burton and Brownie (2006) found 16% and 5% F_1 high-parent heterosis for yield. F_2 bulks have also been tested in standard yield plots along with parents. In four such experiments average mid-parent heterosis was found to be 8%, 11%, 9%, and 7% (Weiss et al., 1947; Loiselle et al., 1990; Gizlice et al. 1993; Manjarrez-Sandoval et al., 1997).

F_1 seed for a standard yield plot was generated in two ways; first by manual pollinations (Brim and Cockerham, 1961; Hillsman and Carter, 1981; Burton and Brownie, 2006) and secondly by insect pollination using genetically male-sterile (ms2ms2 or ms6ms6) female parents (Nelson and Bernard, 1984; Lewers et al., 1998).

The standard measure of the degree of inbreeding of an individual is Wright's coefficient of inbreeding, f (Wright 1921). Originally expressed as the correlation of additive genetic values between two uniting gametes, Wright's coefficient of inbreeding is

now usually interpreted as the probability of identity-by-descent of two alleles at a locus in an individual (Cotterman 1940, Crow 1970, Malecot 1948)

Significant inbreeding depression has been observed where inbred generation bulks have been yield tested together in the same field experiment (Weiss et al. 1943; Brim and Cockerham, 1961; Lewers et al., 1998; Burton and Brownie, 2006; Rahangdale and Raut, 2002). Brim and Cockerham (1961) regressed generation means on percentage inbreeding and found an average linear decline of -6.64 kg/ha^{-1} and -5.59 kg/ha^{-1}. Additionally Burton and Brownie (2006) found an average linear decline of -5.73 kg/ha^{-1} and -2.70 kg/ha^{-1}. Deviations from linearity were non-significant in both experiments.

These observations of heterosis and inbreeding depression are strong evidence that dominance genetic effects for seed yield are probably important in progeny of many cross combinations. Efforts to relate yield heterosis to parental genetic distance or coancestry have had mixed results (Nelson and Bernard, 1984; Gizlice et al. 1993; Cerna et al., 1997; Manjarrez-Sandoval et al., 1997) No significant correlation has been observed between heterosis, specific combination ability or grain yield with genetic distance assessed by AFLP and SSR. (Paterniani et al., 2008)

Table 1.1 Average yield heterosis expressed as a percent of the midparent and/or as a percent of the high parent (Burton, 2008).

	F_1 Yield of spaced plants[1]	F_1 yield in row plots[2]	F_1 yield in single rows[3]	Loiselle et al. (1990)[4]	Gizlice et al. (1993)[5]	Manjarrez-Sandoval et al (1997)[6]	Lewers et al. (1998)[7]	Burton & Brownie (2006)[8]
Mean % Midparent Heterosis	24.9	9.6	48.2	10.8	9.3	6.8	5.0	---
Mean % High parent Heterosis	13.4	4.5	---	---	---	3.1	---	10.5
%F_1's > midparent	78.1	93.6	---	---	---	---	83	100
%F_1's > high parent	54.6	68.1	---	---	---	---	---	100

1. Average of results reported in 9 experiments, 260 F_1's (Burton, 1987; Mehta et al., 1984; Kunta, et al., 1985; Dayde et al., 1989)

2. Average of results reported in 3 experiments, 47 different F_1's (Burton 1987)

3. Average results of 2 experiments, 24 F_1's single rows, 1 yr., 1 location, 3 replications, per plant yield reported (Chauhan, & Singh, 1982; Rahangdale and Raut, 2002)

4. 55 F_2 bulks, 4-row yield plots, 3 replications, 3 locations, 1 year

5. 10 F_2 bulks, 3-row yield plots, 3 replications, 2 locations, 2 years

6. 24 F_2 bulks, 3-row yield plots, 8 replications, 2 locations, 1 year

7. 36 F_1's, 3-row yield plots, 3 replications, 2 years

8. 2 F_1's, 3-row yield plots, 3 replications, 3 locations, 2 years

If heterosis can be established as a predictive tool, breeding methods could be changed to benefit from it. Marker analysis could be done on a population from a cross showing heterosis to determine the genetic basis of heterosis. If heterosis is a result of beneficial dominant loci, then selecting the individuals in the F_2 generation that contains the largest number of beneficial dominant alleles would be advantageous in inbred line development. Direct testing of F_1s for heterosis may not be useful due to the difficulty in producing enough seed for testing, but F_2 bulks could be tested. If the F_2 bulks are predictive of inbred line performance, intercross selection with yield tests of F_2 bulks would be advantageous.

REFERENCES

Brim, C.A., and Cockerham, C.C., 1961. The inheritance of quantitative characters in soybeans. Crop Sci. 1:187-190.

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

Burton, J.W., 1987. Quantitative genetics: Results Relevant to soybean breeding. In: WILCOX, Journal R. (ed.) Soybeans: improvement, production and uses. 2nd ed. Agronomy Monograph, 16, 211-247. ASA-CSSA. SSSA, Madison WI, USA.

Burton, J.W. and Brownie, C., 2006. Heterosis and Inbreeding Depression in Two Soybean Single Crosses. Crop Science 46: (6), 2643-2648.

Burton, J.W., 2008. A review of heterosis in Soybean. Unpublished manuscript.

Chauhan,V.S., Singh, B.B., 1982. Heterosis and genetic variability in relation to genetic divergence in soybean. Indian Journal of Genetics and Plant Breeding, 42, 324-328.

Cerna, F.J., Cianzio, S.R., Rafalski, A., Tingey, S., Dyer, D., 1997. Relationship between seed yield heterosis and molecular marker heterozygosity in soybean. Theoretical and Applied Genetics, 95, 460-467.

Comstock, R.E., and Robinson, H.F., 1952. Estimation of the average dominance of genes. P. 494-516. in: J.F. Gowen (ed.). Heterosis. Iowa State College Press, Ames.

Collins, G.N., 1921. Dominance and the vigor of first generation hybrids. Am. Nat. 55:116–133.

Cotterman, C., 1940. A Calculus for Statistico-Genetics. PhD thesis. Ohio State Univ., Columbus.

Crow, J.F., Kimura, M., 1970. An Introduction to Population Genetics Theory. New York: Harper & Row.

Crow, J. F., 1998. 90 year ago: the beginning of hybrid maize. Genetics 148:923-928.

Crow, J.F. 1999. Dominance and overdominance. p. 49–58. *In* J.G. Coors and S. Pandey (ed.) Genetics and exploitation of heterosis in crops. ASA, CSSA, and SSSA, Madison, WI.

Crow, J.F., 2000. The Rise and Fall of Overdominance. Plant Breeding Reviews, 17: 225-257.

Dayde, J., Ecochard, R., Marmey, P., 1989. The possible influence of cytoplasm on the

performance of reciprocal soybean hybrids. Euphytica, 44, 49-53.

Fehr W.R., 1987. Principles of cultivar development. New York: Macmillan. pp.536

Fisher, R. A., 1922. On the dominance ratio. Proc. Roy. Soc. Edinb. 52:321-341.

Fu, H. and Dooner, H.K., 2002. Intraspecific violation of genetic colinearity and its implications in maize. Proc. Natl. Acad. Sci. USA 99, 9573-9578.

Gardner, C.O., 1963. Estimates of genetic parameters in cross-fertilizing plants and their implications in plant breeding. P. 225-252. in: W.D. Hanson and H.G. Robinson (eds.), Statistical genetics and plant breeding. Special Pub. 982, NAS-NRC, Washington, D.C.

Gizlice, Z., Carter, T.E., Burton, J.W., 1993. Genetic diversity in North American soybean: II. Prediction of heterosis in F2 Populations of southern founding stock using genetic similarity measures. Crop Science, 33, 620-626.

Gowen J.W., 1952. Hybrid vigour in Drosophila. In: Gowen JW, ed. Heterosis. New York: Hafner Publishing Co. 474-493.

Guo, M. et al., 2006. Genome-wide transcript analysis of maize hybrids: allelic additive gene expression and yield heterosis. Theor. Appl. Genet. 113, 831-845.

Hillsman, K.J., and Carter, H.W., 1981. Performance of F1 hybrid soybeans in replicated row trials. Agronomy Abstracts. American Society of Agronomy, Madison, WI, p. 63.

Hull, F.H., 1945. Recurrent selection for specific combining ability in corn. J. Am. Soc. Agron. 37:134–145.

Husband, B.C., Schemske, D.W., 1996. Evolution of the magnitude and timing of inbreeding depression in plants. Evolution, 50, 54-70.

Jones, D.F., 1918. The effects of inbreeding and crossbreeding upon development. Bull.Conn.agric. Exp. Sta. 207: 100.

Kimura, M., 1965. Attainment of quasi linkage equilibrium when gene frequencies are changing by natural selection. Genetics 52:875-890.

Kunta, T., Edwards, L.H., McNew, R.W., Dinkins, R., 1985. Heterotic performance and combining ability in soybeans. Soybean Genetics Newsletter, 12, 97-99.

Lee, J.M., Bush, A., Specht, J.E., and Shoemaker, R., 1999. Mapping duplicate genes in soybean. Genome, 42: 829-836.

Lee, J.M., Grant, D., Ballejos, C.E., and Shoemaker, R., 2001. Genome organization in dicots. II. Arabidopsis as a 'bridging species' to resolve genome evolution events among legumes. Theor. Appl. Genet. 103: 765-773.

Lewers, K.S., St. Martin, S.K., Hedges, B.R., Palmer, R.G., 1998. Testcross evaluation of soybean germplasm. Crop Science, 38, 1143-1149.

Lui and White., 1992. Oxidative stability of soybean oils with altered fatty acid composition. J. Am. Oil Chem. Soc., 69:528-532.

Loiselle, F., Voldeng, H.D., Turcotte, P., St. Pierre, C.A., 1990. Analysis of agronomic characters for an eleven parent diallel of early-maturing soybean genotypes in eastern Canada. Canadian Journal of Plant Science, 70, 107-115.

Malecot, G., 1948. Les Mathematiques de l'Heredite. Paris:Masson

Manjarrez-Sandoval, P., Carter, T.T. Jr., Web, D.M., Burton, J.W., 1997. Heterosis in soybean and its prediction by genetic simliarity measures. Crop Science, 37, 1443-1452.

Mehta, S.K., Lal, M.S. and Beohar, A.B.L., 1984. Heterosis in soybean crosses. IndianJournal of Agricultural Science, 54, 682-684.

Melchinger, A.E., 1999 Genetic diversity and heterosis. In the *Genetics and Exploitation of Heterosis in Crops* (Coors, J.G. and Pandey, S., eds), pp. 99-118, American Society of Agronomy, Inc and Crops Science Society of America, Inc.

Milborrow, B., 1998. A biochemical mechanism for hybrid vigour. J. Exp. Bot. 49: 1063-1071.

Nelson, R. L. and Bernard, R. L., 1984. Production and performance of hybrid soybeans. Crop Science, 24: (3), 549-553.

Paterniani M, Guimarães P, Lüders R, et al. Combining ability, genetic divergence among maize lines and correlation with heterosis. / Capacidade combinatória, divergência genética entre linhagens de milho e correlação com heterose. *Bragantia* [serial online]. 2008;67(3):639-648. Available from: CAB Abstracts, Ipswich, MA. Accessed September 11, 2009.

Rahangdale, S.R., Raut, V.M., 2002. Heterosis and inbreeding depression in soybean (*Glycine max*). Indian Journal of Agricultural Research, 43, 135-144.

Schlueter, J.A., Dixon, P., Granger, C., Grant, D., Doyl, J.J., Shoemaker, R.C., 2004. Mining EST databases to resolve evolutionary events in major crop species. Genome, 47, 868-876.

Soy Stats®, A Reference Guide To Important Soybean Facts & Figures. (2009). Retrieved April 25, 2009, from Soy Stats 2009. Website: http://www.soystats.com/2009/Default-frames.htm

Shoemaker, R., Polzon, K., Labate, J., Specht, J., Brummer, E.C., Olson, T., Young, N., Concibido, v., Wilcox, J., Tamulonis, J., Kochert, G., and Boerma, H.R., 1996. Genome

duplication in soybean (*Glycine* subgenus *soja*). Genetics, 144: 329-338.

Springer, N. M., and R. M. Stupar. 2007. Allelic variation and heterosis in maize: how do two halves make more than a whole? Genome Research 17, no. 3: 264-275.

Stuber C.W., 1994. Heterosis in Plant Breeding. *Plant Breeding Reviews*, 12: 227-251.

Stupar, R.M. and Springer, N.M., 2006. Cis-transcriptional variation in maize inbred lines B73 and Mo17 leads to additive expression patterns in the F1 hybrid. Genetics 173, 2199-2210.

Swanson-Wagner R.A., Jia, Y., DeCook, R., Borsuk, L.A., Nettleton, D., Schnable, P.S., 2006. All possible modes of gene action are observed in a global comparison of gene expression in a maize F1 hybrid and its inbred parents. PNAS, 103, 6805-6810.

Uzarowska, A. et al., 2007. Comparative expression profiling in meristmes of inbred-hybrid triplets of maize based on morphological investigations of heterosis for plant height. Plant Mol. Biol. 63, 21-34.

Wittkopp, P.J. et al., 2004. Evolutionary changes in cis and trans gene regulation. Nature 430, 85-88.

Weiss, M.G., Weber, C.R., Kalton, R.R.. 1947. Early generation testing in soybeans. Journal of the American Society of Agronomy, 39, 791-811.

Wright, S., 1921. Systems of mating. Genetics 6:111-178.

CHAPTER 1

Introduction

Heterosis in soybeans is generally not considered to be of importance. The majority of genetic variance in soybeans is thought to be additive variance (Burton, 1987). With inbreeding, the variance among family means will increase while variance within families will decrease. Based on the assumption of additive variance, and the effects of inbreeding on the inter and intra line variances, breeding methods such as the modified pedigree selection (single seed descent) were developed (Brim 1966).

Glycine max is an autogamous species. As a result autogamy, small flower morphology, and lack of abundant windblown pollen, crossing to develop hybrid seeds is a difficult and time consuming task. To develop a new variety, the usual approach is to cross between two selected cultivars, or elite lines, followed by years of rapid inbreeding using single seed descent (Brim, 1966) until the F_3, F_4, F_5 or F_6 generation. It is only after inbreeding that advanced lines are developed and tested. All commercially available cultivars are currently inbred lines derived by single plant selection in the F_3 or later generations. Because of the common acceptance of inbred lines and the difficulty in producing F_1 seed, hybrid breeding has been neglected. For this reason, the F_1 generation has rarely been compared to succeeding generations.

Despite the overall lack of interest in hybrid breeding, a few F_1 hybrids have been yield tested in standard replicated row plots. Brim and Cockerham (1961) tested F_1s from

two crosses and found an average of 20% high-parent heterosis. Hillsman and Carter (1981) tested F_1s from eight crosses and found an average 6.2% high-parent heterosis. Nelson and Bernard (1984) reported an average 3.3% high parent-heterosis among 37 crosses. Over these 3 experiments, 46 crosses were tested. Of these 46 crosses, 32 exhibited high-parent heterosis.

More recently, Lewers et al. (1998) investigated 36 F_1 test crosses. They used 3 isolines of Clark and Harosoy as female parents and crossed these with 6 genetically diverse soybean lines as their male parents. Clark matures with MG IV cultivars and Harosoy matures with MG II cultivars. The 36 hybrids exhibited an average F_1 mid-parent heterosis for yield between 9.3% and 2.5%. Additionally, Burton and Brownie (2006) found 16% and 5% F_1 high-parent heterosis for yield in two crosses.

Significant inbreeding depression has been observed in soybean in row plots (Weiss et al. 1947; Brim and Cockerham, 1961; Lewers et al., 1998; Burton and Brownie, 2006; Rahangdale and Raut, 2002). Brim and Cockerham (1961) regressed generation means on percentage inbreeding and found an average linear decline of -6.64 kg/ha^{-1} and -5.59 kg/ha^{-1}. Additionally, Burton and Brownie (2006) found an average linear decline of -5.73 kg/ha^{-1} and -2.70 kg/ha^{-1}. Taken together, observations of both heterosis and inbreeding depression are strong evidence that dominance genetic effects for seed yield are probably important in non-inbred progeny of many cross combinations.

The purpose of this experiment was to test four hybrid combinations for heterosis and inbreeding depression with respect to various quantitative traits. One of these was

Hutcheson x Holladay; one of the two hybrids previously tested by Burton and Brownie (2006) which had the highest heterosis. The combination of heterosis and inbreeding depression estimates will provide additional information regarding the presence of dominance in soybeans and its effect upon seed yield and other traits.

Materials and Methods

In this experiment, F_1, F_2, F_3, and F_4 seed were generated for 4 crosses. The F_1 seeds were generated, during the summers of 2004 through 2006. During each of these three crossing seasons, 400 attempts per cross combination were made with the goal to generate enough seed for 2 years of testing in replicated field plots. With a success rate below 50%, an average of less than 200 F_1 seeds was generated for each cross each year. After 4 years of crossing, enough F_1 seed had been generated to complete the first year of hybrid testing at two locations. F_2 seed for this trial were generated in the winter of 2005 in a winter nursery in Puerto Rico. The F_3 seed were generated in the summer of 2006 in Clayton, NC. The F_4 seed were generated in the winter of 2006 in Puerto Rico.

The four crosses used were Hutcheson x Holladay, Satelite x Holladay, 5601T x N00-370, and G98-1053 x N99-244 (Table 2.1). The cultivars used in this study were 'Holladay' (Burton et al., 1996), 'Hutcheson' (Buss et al., 1988), 'Satelite' a cultivar derived from F_3 plant selections from the cross 'Soyola' (Burton et al., 2004) X {'Brim' (Burton et al., 1994) X [N88-431 X (N90-2013 X C1726)]}, '5601T' (Pantalone et al., 2003), 'G98-1053' is an F_5 derived breeding line from the cross of 'Boggs' (Boerma et al., 2000) and 'Doles' (Boerma et al., 1994), 'N00-370' is an F_5 derived breeding line with 'Haskell' (Boerma et al., 1994) as a maternal grandparent and 'Brim' (Burton et al., 1994) as a paternal grandparent, and 'N99-244' an F_5 derived breeding line with 'Hutcheson' as a maternal grandparent and 'Brim' as a paternal grandparent.

Table 2.1 Coefficients of Parentage of the parents used for crossing.

♀ Parent	♂ Parent	Coefficient of Parentage†
Hutcheson	Holladay	0.1470
Satelite	Holladay	0.1029
5601T	N00-370	0.1824
G98-1053	N99-244	0.2023

†A measure of genetic distance between two individuals

In the summer of 2007, the field trial was grown at both Clinton and Plymouth, NC in a randomized complete block design with three replications. The soil in Clinton was Norfolk series, Thermic Typic Kandiudults; the soil in Plymouth was Portsmouth series, Typic Umbraquult. The yield plots consisted of three rows, 4.9 meters long by 0.90 meters wide; 4 meters were harvested from the middle row. Due to a limited amount of F_1 seed, the border rows of the F_1 yield plots were planted in a 50:50 mix of their parents. Bordering the F_1 yield plots with a mix of their parents should have minimized any border effects on F_1 performance due to inter-genotypic competition (Schutz and Brim 1967). The F_1 rows were rogued for selfs based upon pubescence color or flower color with flowers and grey pubescence being recessive traits. A very low percentage of selfs were found. The other generations (F_2 through F_4) had enough seed to plant all three rows with no substitutions. The yield trial in Clinton in 2007 was dropped due to severe drought stress resulting in a very poor stand and inconsistent data. The yield trial at Plymouth was harvested in the fall using a single row research combine.

This experiment was repeated in 2008. The F_1 planting seed were a remnant from the crossing efforts of 2004-2006. The bulked seed from the F_1, F_2, and F_3 rows of the 2007 experiment were sampled and used for the F_2, F_3, and F_4 rows of the 2008 experiment respectively. The experiment was grown at Clinton and Clayton, NC. In Clayton there were two replications, while in Clinton only one replication was grown due to a shortage of F_1 seed.

At each location, flowering date was noted in all plots. The maturity date was scored by dating the number of days past August 1st, and plant height and lodging at maturity were also recorded. After harvest, total seed weight was determined for each plot and in 2008, seed size (grams per 100 seeds) was also determined

Statistical Analysis

The phenotypic data were then analyzed using the MIXED Procedure with a mixed model analysis of variance (ANOVA) in SAS (SAS Institute, 2004). Even though two different years of testing were done, location effects were used instead of years because no location was represented in both years. The phenotypic data was combined over all locations in both years and analyzed. In the combined analysis genotype by location was significant. Because genotype by location was highly significant, the locations from the two years were analyzed separately by year. In the mixed model genotype was treated as fixed and locations, replications within locations, and genotype by location were treated as random. Kenward Roger containment method was used for denominator degrees of freedom.

Contrasts were set up to test each F_1 hybrid against its parent with the highest value for each trait each year. Mid-parent F_1 heterosis was tested in all crosses, by contrasting the F_1 hybrid with the average of its two parents for each trait each year. Since the seed size data was only available from 2008, the year and genotype by year effects were omitted from the mixed model of seed size. The LSDs for the contrasts were estimated using the complete analysis residual because the errors were assumed to be constant across the experiment.

To test for a linear relationship between yield and inbreeding, a regression using PROC REG of SAS was performed on the yield data separated by cross and year. The inbreeding coefficient was coded as 0, 0.5, 0.75, and 0.875 for the F_1, F_2, F_3, and F_4 respectively. This analysis was repeated for additional traits. The coefficient of parentage for each cross combination was regressed on yield means by year and tested for at significance level 0.05.

Results

Tests for mid-parent heterosis were conducted for each year and cross combination. G98-1053 x N99-244 exhibited non-significant high- and mid-parent F_1 yield heterosis in 2007 (0.4% and 5.4% respectively), but had significant high-parent F_1 yield heterosis in 2008 of 21.7%. 5601T x N00-370 exhibited non significant high-parent F_1 yield heterosis of 11.1% with significant mid-parent F_1 yield heterosis of 20.9% in 2007. In 2008 5601T x N00-370 had significant high-parent F_1 yield heterosis of 30.7%. Satelite x Holladay exhibited non significant high-parent and mid-parent F_1 yield heterosis of 3.0% and 3.7% respectively in 2007. In 2008 Satelite x Holladay exhibited non significant high-parent F_1 yield heterosis of 12.5% with significant mid-parent F_1 yield heterosis of 22.7%. Hutcheson x Holladay exhibited non significant high-parent F_1 yield heterosis of 23.9% with significant mid-parent F_1 yield heterosis of 24.0% in 2007. In 2008 Hutcheson x Holladay non significant high- and mid-parent F_1 yield heterosis of 2.5% and 7.1% respectively. (Table 2.2 and Table 2.3)

Table 2.2. Yield means of parents, F_1, and inbred generations from all four crosses averaged over both locations in 2007.§

Generation	G98-1053 x N99-244	5601T x N00-370	Satelite x Holladay	Hutcheson x Holladay
	----------kg ha^{-1}----------			
♂ Parent¥	2733	3063	2867	2941
♀ Parent	3013	2565	2937	2937
Mid-Parent	2873	2814	2915	2939
F_1	3027	3400	3023	3643
F_2	2938	2599	3133	3446
F_3	2857	2561	2649	2953
F_4	2639	2579	2447	3428
LSD 0.05†	715	715	715	715
LSD 0.05‡	619	619	619	619

§Three replications at Plymouth.

†Least Significant Difference between parents and generations

‡Least Significant Difference between mid-parent and generations

¥Female parent listed first in cross.

Table 2.3. Yield means of parents, F_1, and inbred generations from all four crosses averaged over both locations in 2008.§

Generation	G98-1053 x N99-244	5601T x N00-370	Satelite x Holladay	Hutcheson x Holladay
	----------kg ha^{-1}----------			
♂ Parent¥	3668	3284	2999	3052
♀ Parent	3955	3238	3722	3722
Mid-Parent	3812	3261	3361	3387
F_1	4797	4314	4027	3519
F_2	3800	3429	3344	2975
F_3	3827	3834	2868	3375
F_4	3666	3328	3284	3266
LSD 0.05†	804	804	804	804
LSD 0.05‡	696	696	696	696

§Two replications at Clayton and one replication in Clinton.

†Least Significant difference between parents and generations

‡Least Significant Difference between mid-parent and generations

¥Female parent listed first in cross.

The lack of consistency for heterosis between years is a result of the significant genotype by location interaction across both years for yield (Appendix 1, 2, 3). Of the five instances of significant mid-parent F_1 heterosis, only three of these had significant high-parent F_1 heterosis. Those crosses that were significant for mid-parent F_1 heterosis in 2007 were not all significant in 2008 (Hutcheson x Holladay). (Appendix 1)

Regarding seed size in 2008, none of the crosses had significant high-parent F_1 heterosis. However, all four had significant mid-parent F_1 heterosis (Table 2.4). 5601T x N00-370 had a high-parent F_1 heterosis of 4.2% with a significant mid-parent F_1 heterosis of 6.1%. G98-1053 x N99-244 had a high-parent F_1 heterosis of 1.0% with a significant mid- parent F_1 heterosis of 14.5%. Hutcheson x Holladay had a high-parent F_1 heterosis of 2.4% with a significant mid-parent F_1 heterosis of 6.6%. Satelite x Holladay had a high-parent F_1 heterosis of 4.6% with a significant mid-parent F_1 heterosis of 5%. (Table 2.4)

Table 2.4. Seed size means of parents, F_1, and inbred generations from all four crosses averaged over both locations in 2008.§

Generation	5601T x N00-370	G98-1053 x N99-244	Hutcheson x Holladay	Satelite x Holladay
	----------grams 100 seeds^{-1}----------			
♂ Parent	15.72	12.90	16.34	17.54
♀ Parent	16.53	16.53	17.81	17.81
Mid-parent	16.13	14.72	17.08	17.68
F_1	17.70	17.10	19.00	18.43
F_2	17.35	14.89	16.93	17.33
F_3	15.82	14.12	16.98	16.27
F_4	15.34	15.88	17.84	16.99
LSD .05†	1.8365	1.8365	1.8365	1.8365
LSD .05‡	1.5903	1.5903	1.5903	1.5903

§Two replications at Clayton and one replication in Clinton.

†Least Significant Difference between parents and generations

‡Least Significant Difference between mid-parent and generations

Both high parent and mid-parent F_1 heterosis were non-significant for lodging, height or maturity date (data not presented).

When yields of each generation are regressed upon the inbreeding coefficient, 5601T x N00-370 showed significant inbreeding depression in 2007 and tended towards

inbreeding depression in 2008. This cross had inbreeding depression in 2007 and 2008 of nearly -10 kg ha^{-1} per 1% inbreeding (-9.8 kg ha^{-1} and -9.3 kg ha^{-1} respectively). Satelite x Holladay had significant inbreeding depression in 2008, with a decline of nearly 11 kg ha^{-1} (-10.9 kg ha^{-1}). The other cross by year combinations had overall declines in yield with inbreeding, though they were not significant.

When mean generation seed size was regressed upon the inbreeding coefficient, all four crosses had declines in seed size (mg per seed) as inbreeding increased. Two crosses, 5601T x N00-370 and G98-1053 x N99-244, had significant inbreeding depression with declines of 24 mg and 23 mg per 1% inbreeding respectively. None of the crosses showed any decline in height due to inbreeding. The trend for all the crosses was increased height with inbreeding, but this was not statistically significant. When lodging score and maturity date were regressed on inbreeding coefficient, none of the crosses showed any significant trends with respect to inbreeding. (Table 2.5)

The regression of percent high-parent F_1 yield heterosis upon Coefficient of Parentage (CP) was non-significant for all traits over all years. The high-parent F_1 heterosis estimates for yield and seed size were separated by year and regressed on CP. The correlation between heterosis and CP was very low.

Table 2.5. Regression of yield upon inbreeding coefficient (F) from all four crosses§.

	Yield in 2007	Yield in 2008	Seed Size	Height	Lodging (score)¥	Maturity (date)
	----------b+----------					
	-------kg ha^{-1}------		mg 100 seed^{-1}	meters		
5601T x N00-370	-9.81*	-9.27*	-24*	0.03	0.002	0.06
Satelite x Holladay	-6.30	-10.85*	-20	0.04	0.003	0.03
Hutcheson x Holladay	-4.87	-2.25	-11	0.03	0.001	-0.03
G98-1053 x N99-244	-3.74	-12.68	-23*	0.07	0.0001	0.01

§Two replications at Clayton and one replication in Clinton.

+ The slope in the regression equation per 0.001 change in F in $Y = a + bF + e$, were F is the inbreeding coefficient.

* Significant at p value 0.05.

¥ Scored on a rating of 1-4 with 4 being the most lodged

Discussion

Heterosis

The results in 2007 and 2008 revealed a significant location by genotype interaction, so that an analysis by year was needed. This interaction was likely due to a very severe drought in 2007 and adequate rainfall in 2008. Without a second location, the location effect in 2007 was confounded with the year effect, thus confounding the genotypic variance estimates with genotype by environment effects. In 2008 only one replication was possible at the Clinton location. Without multiple replications at a location, poorer yielding areas of the field could result in aberrant data.

The results do show significant heterosis in both years for both high-parent and mid- parent. However, the lack of consistency across years could mean two things. First the unusual conditions experienced during the drought of 2007 may have introduced a significant, albeit rare, genotype by environment interaction. Secondly the data suggest that there is an underlying instability in heterosis in soybeans that was dramatically exposed due to extreme weather conditions. This instability of heterosis would reduce its value to the breeder. If heterosis is magnified to a significant level only through its interaction with a particular set of environmental conditions, then it would be based less upon a set of favorable genes than upon a favorable environment. If a cross were selected based on heterotic performance of the F_2, then inbred lines developed in succeeding generations would have favorable gene combinations but might not have stability across environments. There were differences in maturity between the parents which would then

lead to expected segregation in the F_2 through F_4 generations which could have contributed to the differences seen between years.

The Hutcheson x Holladay cross did show a significant 23.8% high parent heterosis in 2007, which was slightly larger than the heterosis found by Burton and Brownie (2006). The heterosis estimate was not observed in 2008, but the cross was numerically larger (2% high-parent heterosis). This suggests that the genotype by environment effects in this study may be unique to this study. This assertion must be tempered with the understanding that while 2007 was the year with the most drastically different environment, it was the year that Hutcheson x Holladay behaved as it had in the Burton and Brownie (2006) study.

In this study, we found no relationship between magnitude of F_1 high-parent heterosis and CP ($r = 0.01$ in 2007 and $r = 0.04$ in 2008). This lack of relationship has been noted in other studies as well (Burton and Brownie, 2006; Manjarrez-Sandoval et al., 1997). Manjarrez-Sandoval et al. (1997) investigated heterosis of F_2 bulks in 24 crosses and found correlation coefficients between yield F_2 heterosis and CP between 0.000001 and 0.04. In addition to using CP, Manjarrez-Sandoval et al. (1997) also estimated genetic distance using restriction fragment length polymorphisms (RFLP). Additional studies attempting to relate yield heterosis to parental genetic distance or coancestry have had mixed results (Nelson and Bernard, 1984; Gizlice et al. 1993; Cerna et al., 1997; Manjarrez- Sandoval et al., 1997). Our results, along with previous research, continue to show a lack of predictive power of CP and genetic distance for heterotic performance.

There are two basic theories explaining heterosis: dominance and overdominance. The dominance hypothesis assumes that, in hybrids, detrimental recessives from one parent are masked by dominant alleles from the other. The overdominance hypothesis assumes that, at least at some loci, heterozygotes are superior to either homozygote. These two hypotheses have been treated as opposites, but they are not mutually exclusive. In fact, both are consistent with inbreeding decline and heterosis, and distinguishing between the two can be difficult (Crow 2000). If heterosis is due to overdominance, then F_1 performance would not be predictive of inbred line performance because overdominance cannot be fixed in inbred lines, whereas dominance can be. This study found significant inbreeding depression when the yields of the four generations were regressed upon F. The inbreeding depression found in this study points towards dominance. Current thought is that the majority of genetic variance in soybeans is additive variance (Burton, 1987). The inbreeding depression found in this study plus the fact that current thought suggests that additive variance is a significant component of genetic variance means there is probably both additive and dominance variance in these four populations. We cannot say much about the relative size of the dominance and additive variance because in the F_2 where you first see segregation, inbreeding depression has already reduced dominance by ½. The variance among progeny means increases and becomes all additive and additive types of epistasis. The increase in additive variance was the foundation for the development of SSD by Brim and is the rationale for disregarding dominance as important in soybean breeding. But dominance may be important if it can be used to predict inbred line performance. This may be a reason for the implementation of different breeding methods.

Inbreeding Depression

As with heterosis, there was a genotype by year effect for inbreeding depression. The same explanations for the instability of the yield values regarding heterosis apply here.

The resulting lack of significance in height, lodging score, and maturity date when regressed on percent inbreeding show that the decrease in yield was not due to changes in plant height, lodging or maturity, but rather due to a decrease in seed size and possible number of seeds produced per hectare.

In the previous studies on heterosis, significant inbreeding depression was observed in cases where inbred generation bulks were yield tested together in the same field experiment (Weiss et al. 1943; Brim and Cockerham, 1961; Lewers et al., 1998; Burton and Brownie, 2006; Rahangdale and Raut, 2002). The inbreeding depression values found in this study fall well within the range of previously reported values. This study, as with previous studies, there was no significant deviation from linearity and thus no evidence for epistasis.

Inbreeding depression has been rarely observed with soybeans because yield is usually not measured until an advanced stage of inbreeding. In this study, as well as in others (Burton and Brownie, 2006; Brim and Cockerham, 1961; Lewers et al., 1998; Rahangdale and Rout, 2002), statistically significant linear inbreeding depression is clear evidence that dominance effects are present.

Finally, in self pollinated crops, the genetic load is assumed to be low. Natural

selection and plant breeding would be expected to eliminate major deleterious alleles. Additionally, favorable dominant alleles would increase in the population over less favorable alleles. Therefore, inbreeding depression would likely not be a result of the exposure of major deleterious alleles but the decrease in dominance in the population. With an increase in F, the lesser yielding allele would become exposed thus decreasing the yield of the population. The F_2 segregates ¼ AA, ½ Aa, ¼ aa, and the F_∞ segregates ½ AA and ½ aa. With dominance, there would be two phenotypic classes; 3/4 A_ and 1/4 aa, this would change in the F_∞ 1/2 AA and 1/2 aa. The change in genotype segregation would mean that there should be a decline in yield from the F_2 to the F_∞. Without dominance, the phenotypic mean values would not change from F_2 to F_∞. There have been several previous studies showing redundancy in the genome with hybridization-based genetic maps (Shoemaker et al. 1996; Lee et al. 1999; Lee et al. 2001) and RFLP probes which detected extensive homoeologous relationships between linkage groups (Shoemaker et al. 1996). This genome duplication would contribute to the conservation of mildly deleterious genes (Husband and Schemske, 1996) which would partly explain heterosis in an F_1 hybrid that combined favorable alleles at both loci, one contributed from each parent.

Burton and Brownie (2006) postulated that heterosis would be evidence of superior gene combinations. If efficient selection among cross combinations in F_2 or F_3 bulk evaluations was done for heterosis, selection may lead to a higher frequency of high-yielding pure lines than breeding methods that ignore dominance. Additionally, knowing the mode of action of the alleles in question would be of interest. As concluded by

Swanson-Wagner et al. (2006) in corn, all modes of action, dominance, additive, and overdominance, were present. Results here show that there is dominance involved in heterosis in soybeans. Testing the soybean genome to determine if dominance verses overdominance is involved in heterosis in soybeans would help in the understanding of the causes and uses of heterosis.

REFERENCES

Brim, C.A., and Cockerham, C.C., 1961. The inheritance of quantitative characters in soybeans. Crop Sci. 1:187-190

Brim, C.A., 1966. A modified pedigree method of selection in soybeans. Crop Sci. 6:220

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

Boerma, H.R., R.S. Hussey, D.V. Phillips, E.D. Wood, and S.L. Finnerty., 1994b. Registration of 'Doles' soybean. Crop Sci. 34:1411.

Boerma, H.R., Hussey, R.S., Philips, D.V., Wood, E.D., Rowan, G.B., and Finnerty, S.L., 2000. Registration of 'Boggs' soybean. Crop Sci. 40:295.

Burton, J.W. 1987. Quantitative genetics: Results Relevant to soybean breeding. P. 211-247. *In*: J.R. Wilcox (ed.) Soybeans: Improvement, production and uses. 2nd ed. ASA Monogr. 16. ASA, CSSA, and SSSA, Madison WI, USA.

Burton, J.W., Carter, Jr. T.E., and Huie., E.B. 1988. Registration of 'Brim' soybean. Crop Sci. 34:301

Burton, J.W., Carter, Jr. T.E., and Huie., E.B. 1994. Registration of 'Brim' soybean. Crop Sci. 34:301

Burton, J.W., Carter, Jr. T.E., and Huie., E.B. 1996. Registration of 'Holladay' soybean. Crop Sci. 36:467

Burton, J.W., 2004. Registration of 'Soyola' soybean. Crop Sci. 44:687-a-688-a.

Burton, J.W. and Brownie, C., 2006. Heterosis and Inbreeding Depression in Two Soybean Single Crosses. Crop Sci. 46:2643-2648.

Buss, G.R., Camper, Jr. H.M., Toane, C.W., 1988. Registration of 'Hutcheson' soybean. Crop Sci. 28:1024

Cerna, F.J., Cianzio, S.R., Rafalski, A., Tingey, S., Dyer, D., 1997. Relationship between seed yield heterosis and molecular marker heterozygosity in soybean. Theoretical and Applied Genetics. 95:460-467.

Crow, J.F., 2000. The Rise and Fall of Overdominance. Plant Breeding Reviews, 17:225 257.

Gizlice, Z., Carter, T.E., Burton, J.W., 1993. Genetic diversity in North American soybean: II. Prediction of heterosis in F2 Populations of southern founding stock using genetic similarity measures. Crop Sci. 33:620-626.

Hillsman, K.J., and Carter, H.W., 1981. Performance of F1 hybrid soybeans in replicated row trials. p. 63 Agronomy Abstracts. ASA, Madison, WI.

Husband, B.C., Schemske, D.W. (1996): Evolution of the magnitude and timing of inbreeding depression in plants. Evolution, 50, 54-70.

Lee, J.M., Bush, A., Specht, J.E., and Shoemaker, R. 1999. Mapping duplicate genes in soybean. Genome, 42: 829-836

Lee, J.M., Grant, D., Ballejos, C.E., and Shoemaker, R. 2001. Genome organization in dicots. II. Arabidopsis as a 'bridging species' to resolve genome evolution events among legumes. Theor. Appl. Genet. 103: 765-773

Lewers, K.S., St. Martin, S.K., Hedges, B.R., Palmer, R.G., 1998. Testcross evaluation of soybean germplasm. Crop Sci. 38:1143-1149.

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

Nelson, R. L. and Bernard, R. L., 1984. Production and performance of hybrid soybeans. Crop Sci. 24:549-553.

Pantalone, V.R., Allen, F.L., and Landau-Ellis, D., 2003. Registration of '5601T' soybean. Crop Sci. 43:1123-1124

Rahangdale, S.R., Raut, V.M., 2002. Heterosis and inbreeding depression in soybean (Glycine max). Indian Journal of Agricultural Research. 43:135-144.

SAS Institute Inc., SAS 9.1.3 Help and Documentation, Cary, NC: SAS Institute Inc., 2000-2004.

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

Schutz, W.M., and Brim C.A., 1967. Inter-Genotypic Competition in soybeans. I. Evaluation of effects and proposed field plot design. Crop Sci. 7:371-376.

Swanson-Wagner R.A., Jia, Y., DeCook, R., Borsuk, L.A., Nettleton, D., Schnable, P.S., 2006. All possible modes of gene action are observed in a global comparison of gene expression in a maize F1 hybrid and its inbred parents. PNAS, 103, 6805-6810.

Weiss, M.G., Weber, C.R., Kalton, R.R., 1947. Early generation testing in soybeans. Journal of the American Society of Agronomy. 39:791-811.

CHAPTER 2

Introduction

Soybean seed is valued for both its protein and oil. In addition to maintaining or increasing seed protein and oil concentrations, crop productivity is an ongoing breeding goal. In 2008, 30% of all vegetable oil and 68% of protein meal consumption in the world was derived from soybean. Because soybean is such an important source of protein and oil, demand for the crop is likely to increase as world population grows. Therefore, increases in soybean yield will continue to be paramount for society.

In soybean breeding, it is currently difficult to accurately predict good cross combinations that lead to improved yield. That is, which potential parents will combine to produce transgressive segregation for particular quantitative traits among their progeny? Breeding programs select a pair of good lines for mating while making sure that they are not closely related, using coefficient of parentage as a measure of relationship. Often two lines are mated with a desire to combine traits the two have individually.

Weiss et al. (1947) suggested that F_1 or F_2 performance might be used to predict inbred line performance in soybean based on the assumption that greater heterosis was indicative of greater genetic distance between parents. However they found that the degree of heterosis found in spaced F_1 plants was of limited predictive value in the crosses studied.

Harrington (1940) tested bulked F_2 lines from six wheat crosses in replicated yield

trials and found heterosis, indicating yield potentialities of segregates from those crosses. Weiss (1947) found in soybeans Harrington (1932) found in one wheat cross that evaluation of several hundred single F_2 plants correctly predicted the value of the cross as to earliness, height, stem rust reaction, and seed characteristics; however the F_2 yield results were misleading.

Immer (1941) found an inconsistency in phenotypic data between F_1 spaced plants and bulked F_2, F_3, and F_4 populations in barley. The six barley crosses Immer used showed substantial differences among crosses that were consistent across the F_2, F_3, and F_4 bulked populations. Lack of agreement was attributed to interaction of crosses x method of planting, particularly noticeable when the F_2, F_3, and F_4 bulked populations were grown in different years.

If the F_2 generation shows heterosis, it could be inferred that the F_1 generation exhibited heterosis as well. If the best crosses can be predicted based upon F_2 performance relative to parents(i.e. having the highest probability of generating high yielding pure lines), the poorer crosses (those with lowest F_2 yields) could be eliminated before deriving and testing inbred lines, thus saving resources (Burton and Brownie 2006; Harrington, 1940).

The first objective of this experiment was to determine whether breeding lines from crosses that showing heterosis for yield have more transgressive segregation than crosses with no heterosis. The second objective was to investigate a possible relationship between heterosis of the F_2 generation of a biparental cross and genetic variation among inbred

lines derived from the cross. Populations derived from crosses with heterosis may have higher progeny variance than those populations without heterosis. With a larger difference between the heterotic hybrid and the mid-parent value a greater amount of segregation is possible. Generating large amounts of F_1 seed adequate for yield testing is difficult. A large number of seed is needed for yield testing in standard row plots and the success rate for crossing is low. Generating the needed seed for all six populations in this study was not possible due to time and labor constraints. Therefore F_2 seed were generated and tested for heterosis.

Materials and Methods

Six crosses were selected (Table 3.1). These crosses were selected based on their having produced a number of lines that were productive in NC preliminary yield trials. The crosses with the largest number or highest proportion of selected lines were assumed to be the best crosses in the year that they were tested. Crosses were selected from multiple years. N90-516 x N92-598 (Cross A) and Dillon (Shipe et al., 1997) x N93-66 (Cross B) were tested in NC preliminary trials in 2001. N89-91 x G93-9009 (Cross F) was tested in NC preliminary trials in 2002. Au94-863 x N96-180 (Cross C), N94-7441 x N90-845 (Cross D), and N94-537 x K1424 (Cross E) were tested in NC preliminary trials in 2004.

N90-516 is an F_5 derived breeding line with 'Hutcheson' (Buss et al., 1988) as the maternal parent and GaSoy 17 and N77-940 as paternal grandparents. N92-598 is an F_5 derived breeding line with 'Johnston' (Burton et al., 1987b) as a maternal grandparent and paternal breeding lines derived from 'Essex' (Smith and Camper, 1973), 'Cenntenial' (Hartwig and Epps, 1977), and 'Lee'. N93-66 is an F_5 derived breeding line with 'Epps' as a maternal grandparent and 'Holladay' (Burton et al., 1996) as a male parent. Au94-863 is an F_5 derived breeding line from a cross of 'Cook' (Boerma et al., 1992) and Au87-727 a breeding line derived from 'Forrest' (Hartwig and Epps, 1973), 'Centennial', 'Tracy' (Hartwig, 1974) and 'Ransom' (Brim and Elledge, 2001). N96-180 is an F_5 derived breeding line from maternal parent 'Cook' and paternal parent N87-298, a breeding line derived from 'Essex', 'Ransom', and 'Lee 68'. N94-7441 is a F_5 derived

breeding line with maternal grandparents ‘GaSoy 17’ and ‘Vance’ and paternal parent ‘Pearl’ (Carter et al. 1995). ‘Vance’ is derived from a cross of ‘Essex’ and an unknown soybean. N90-845 is an F_5 derived breeding line from a cross of two breeding lines. These lines have ‘Young’ (Burton et al., 1987a), ‘Tracy’, ‘Ransom’, and ‘Lee’ in their pedigree. N94-537 is an F_5 derived breeding line from the cross of ‘Cook’ and N87-325. N87-325 is an F_5 derived breeding line developed from ‘Essex’, ‘Ransom’, and ‘Lee’. K1424 is an F_5 derived breeding line developed from the cross of ‘Hutcheson’ and Asgrow A4715. Asgrow A4715 is a breeding line developed from ‘Forrest’, ‘Tracy’, ‘Essex’, and ‘Douglas’ (Nickell et al., 1982). N89-91 is an F5 derived breeding line from the cross of N80-50232 and N81-320. N80-50232 is a breeding line derived from ‘Forrest’ and Bragg’ (Hinson and Hartwig, 1964). N81-320 is a breeding line derived from ‘Lee’, ‘Tracy’, and ‘Ransom’. G93-9009 is an F5 derived breeding line developed from ‘Bedford’ (Hartwig and Epps, 1978), ‘Forrest’, ‘Centennial’, and ‘Picket 71’.

Table 3.1 Parents and their Coefficients of Parentage for each cross

Cross	Made	♀ Parent	Maturity Group	♂ Parent	Maturity Group	Coefficient of Parentage†
A	1996	N90-516	V	N92-598	VI	0.2766
B	1996	Dillon	VI	N93-66	V	0.2172
C	1999	Au94-863	VIII	N96-180	V	0.4398
D	1999	N94-7441	VII	N90-845	VII	0.1730
E	1999	N94-537	VIII	K1424	II	0.2071
F	1997	N89-91	VI	G93-9009	VI	0.3337

†A measure of genetic distance between two individuals

The $F_{6:7}$ generations were generated from F_6 seed kept in cold storage. The F_6 seed had been generated by single seed descent from the F_2 generation. These seed were planted at Clayton, NC and single plant harvested to form $F_{6:7}$ lines. Seeds from 75 $F_{6:7}$ lines from each population were grown in rows at Clayton, NC to increase the seed for testing in 2007. The $F_{6:8}$ lines were then used for yield testing in 2008. 37 inbred lines within each population were selected from the larger number group of 75 based upon their similarity in maturity. This selection reduced the confounding factor that maturity can introduce when comparing line yield performance. Parents of each population were mated in 2006 and 2007 to regenerate the original F_1s. The F_1s were grown in a winter nursery in Puerto Rico to generate the F_2 seed that was used for yield testing in 2008.

The parents, the F_2 bulks, and 37 $F_{6:8}$ lines were then yield tested in replicated field experiments. The yield trials were carried out at three locations, Clinton, Kinston, and

Plymouth, NC, in the summer of 2008. A nested split plot experimental design was used. The split plot contained 6 whole plots divided into 40 subplots. A cross was assigned to each whole plot. The $F_{2,}$ the two parents, and 37 $F_{6:8}$ derived lines were randomly assigned to the 40 subplots in each whole plot.

The plots were planted using a three row research planter. The rows were 0.97 m wide and 5.8 m long and were end trimmed before harvest to establish a uniform row length. The plots were trimmed to a length of 4.9 m at Clinton and Plymouth and a length of 4.3 m at Kinston. Only the center row was harvested for yield.

At maturity, plant height and lodging were recorded. After harvest, total seed weight and size (grams per 100) were determined for each plot. All plot yields were converted to kilograms per hectare.

Data were analyzed using the MIXED procedure in SAS (SAS Institute, 2004). A mixed model was developed to analyze the phenotypic data. In the mixed model genotype was treated as fixed and locations, replications within locations, and genotype by location were treated as random. The analysis was combined over locations. Height, yield, and seed size data were analyzed by population. Type III Sums of Squares were obtained in PROC MIXED and used to estimate genetic and phenotypic variance for each of the $F_{6:8}$ populations. Kenward Roger containment method was used for denominator degrees of freedom. The estimated genetic and phenotypic variances were used to calculate heratibility using the method described by Hallauer and Miranda (Hallauer and Miranda, 1981). Finally the correlation between percent mid-parent heterosis and heritability,

genetic variance and Genetic Coefficient of Variance (GCV) were estimated. GCV was estimated by taking the square root of genetic variance and dividing by the population mean.

Results

Yield

The analysis of variance of yield for each population of inbred lines show there is significant genotypic variance in populations A, C, and E. (Table 3.2).

The F_2 from each population was tested against its parents to determine if it showed any high-parent or mid-parent heterosis for yield. None of the six populations had significant high-parent or mid-parent heterosis. ($p > 0.05$). Population A had a mid-parent heterosis of 15% at p value 0.06 (LSD is 423.95). The yield means of the ♀ Parent (N94-537) and ♂ Parent (K1424) of population E were significantly different. (Table 3.3)

The highest and lowest $F_{6:8}$ lines were contrasted against the parents. Population A was the only population which exhibited heterosis for yield. With the trend towards mid-parent heterosis of population A, it also had two significant ($p = 0.05$) transgressive segregates with a yield range of 3342 to 3451 kg ha^{-1}. Two additional inbred lines were significant at p values 0.074 and 0.0674 with a yield range of 3250 to 3322 kg ha^{-1}. The other populations exhibited no heterosis as well as no transgressive segregation.

Table 3.2 Yield (kg ha^{-1}) Means Squares for each population tested at three NC locations in 2008.

Source	Df	A	B	C	D	E	F
Locations	2	5222826	985886*	5861624	1918412*	7698254	12500018*
Rep(Locations)	3	1464049	1080002	1927158	209463	1285877	417334
Lines	36	727377*	230031	253417*	420595	388665*	175805
Locations x Lines	77	210616	220185*	149020	412198*	219092*	136143*
Residual	82	191986	129084	152369	149051	138056	83325

*Significant at significance level 0.05.

Table 3.3 Yield Means (kg ha^{-1}) for each population $F_{6:8}$ lines, F_2 bulks, and parent tested at three NC locations in 2008.

	A	B	C	D	E	F
♀ Parent	2662.95	3176.82	3347.23	3129.43	3513.05	2595.11
♂ Parent	2989.08	2826.44	3207.90	2963.83	2863.64	2978.63
Mid-Parent	2826.02	3001.63	3277.57	3046.63	3188.35	2786.87
F_2	3250.00	3121.57	3143.45	3276.77	3220.58	2947.84
$F_{6:8}$†	2841.09	2803.49	3397.36	3086.73	2981.74	2894.35
LSD .05‡	578.36	550.79	468.14	762.20	548.70	468.93
LSD .05£	501.42	477.00	405.42	682.91	475.19	419.70

†Average of 37 $F_{6:8}$ lines

‡ For the difference between parents and inbred lines

£ For the difference between F_2 and Mid-parent

Observed heritability ranged from as low as 0.04 to as high as 0.71. The highest heritability was observed population A which had F_2 heterosis and transegregating lines. Populations C and E also had significant genetic variance along with h^2 of 0.41 and 0.44 respectively. (Table 3.4) Populations B, D, and F had the lowest heritabilities, which was expected as they also had non-significant genotypic variance. The correlation between yield mid-parent heterosis and heritability, variance and GCV were all non significant. The correlation for mid-parent heterosis and genetic variance was 0.592 with significance level 0.22.

Table 3.4 Genetic variance, Phenotypic variance, and Heritability of yield by population.

Population	GCV	Genetic Variance	Phenotypic Variance	h^2	Standard Error of h^2
A	0.103	86127	121230	0.71	0.31
B	0.014	1641	38339	0.04	0.71
C	0.039	17400	42236	0.41	0.47
D	0.012	1400	70099	0.02	0.72
E	0.056	28262	64778	0.44	0.46
F	0.028	6610	29301	0.23	0.59

Seed Size

Analysis of variance of seed size for each population of inbred lines showed that there is significant variation within each population. (Table 3.5)

Populations A, B and C had significant mid-parent heterosis at significance level 0.05. The parents for Populations A, C, D, and F were significantly different. Populations D and E have numerical mid parent heterosis; however, this heterosis is not statistically significant. (Table 3.6)

Heritability estimates for seed size were the highest of the three traits, ranging from 0.58 up to 0.91. (Table 3.7)

Table 3.5 Seed size (grams per 100 seed) Means Squares for each population tested at three NC locations in 2008.

Source	Df	A	B	C	D	E	G
Locations	2	0.00	0.16	10.38*	5.26	0.89	19.87
Rep(Location)	2	4.15	7.52	0.64	2.87	7.83	6.15
Lines	36	7.08*	1.40*	11.11*	3.14*	5.74*	4.93*
Lines x Locations	35	0.61	0.59	1.04*	0.87*	1.46*	0.54*
Residual	57	0.46	0.64	0.62	0.28	0.60	0.15

*Significant at significance level 0.05.

Table 3.6 Seed Size means in grams per 100 seeds for parents, $F_{2,}$ and $F_{6:8}$ lines by population over locations.†

	A	B	C	D	E	F
♀ Parent	16.03	15.85	16.63	8.70	15.75	17.23
♂ Parent	16.87	17.03	18.18	13.55	15.25	12.25
Mid-Parent	16.45	16.44	17.41	11.125	15.50	14.74
F_2	17.78	17.58	18.48	11.7	16.05	14.20
$F_{6:8}$	15.34	13.28	18.00	11.19	15.08	12.95
LSD 0.05‡	1.250	1.034	1.456	1.321	1.736	1.315
LSD 0.05£	1.043	0.896	1.261	1.144	1.503	1.201

†Two replications in two locations

‡ For the difference between the two entries

£ For the difference between F_2 and Mid-parent

Four of the six populations had transegregating lines for seed size. Population A has six lines which are significantly greater than the highest parent. Population B had two lines which are greater than the highest parent. Population C had five lines which are greater than the highest parent. Finally, Population E, which did not have mid-parent F_2 heterosis, has five transegregating lines. (Table 3.8) The correlation between seed size mid-parent heterosis and heritability, variance and GCV were all non significant.

Table 3.7 Genetic variance, phenotypic variance, and heritability of seed size for each population over two locations.†

Population	GCV	Genetic Variance	Phenotypic Variance	h^2	Standard Error of h^2
A	0.068	1.08	1.18	0.91	0.24
B	0.028	0.14	0.23	0.58	0.37
C	0.072	1.68	1.85	0.91	0.24
D	0.055	0.38	0.52	0.72	0.30
E	0.056	0.71	0.96	0.75	0.29
F	0.066	0.73	0.82	0.89	0.24

†Two replications per location

Table 3.8 Mid-parent and range of inbred line seed size means (grams per 100 seed) for lines significantly greater than the highest parent averaged over two locations†.

Population	Midparent	Number	Transegregating lines
			Range
A	16.47	6	17.5 - 19.75
B	16.44	2	17.56 - 19.01
C	17.41	5	19.45 - 20.56
E	15.50	5	17.17 - 18.73

†Two replications per location

Height

The analysis of variance of plant height for each population of inbred lines showed that there is significant genotypic variation within all populations except population D. (Table 3.9)

Table 3.9 Height Means Squares for each population tested at three NC locations in 2008.

Source	Df	A	B	C	D	E	G
Location	1	529.73	105.57	425.68	259.57	1552.54	550.74
Rep(Location)	2	3.92	559.17	137.95	23.2	361.55	343.75
Lines	36	36.93*	25.80*	36.29*	13.45	54.88*	53.46*
Location x Lines	36	16.60	15.98	26.49	11.18	34.78*	16.27
Residual	72	17.95	13.64	20.87	12.13	18.40	23.19

*Significant at significance level 0.05.

The height data was also analyzed in the same manner as seed weight and yield. However there was no statistically significant heterosis or transgressive segregating lines found. All of the populations except C have F_2 height values that are numerically larger than the midparent value, however none of these are significant at 0.05 p value (Table 3.10).

Table 3.10 Height means for each Population in centimeters averaged over locations.

Generation	A	B	C	D	E	F
♀ Parent	26.25	37.12	40.25	37.50	44.50	44.75
♂ Parent	31.25	27.75	40.25	31.25	35.50	39.75
Mid-Parent	28.75	32.44	40.25	34.38	40.00	42.25
F_2	32.25	38.25	37.75	39.00	42.00	43.76
$F_{6:8}$†	28.30	36.18	39.32	35.77	38.56	40.11
LSD 0.05‡	5.626	6.488	7.437	4.616	8.548	5.785
LSD 0.05£	4.873	5.380	6.440	3.998	7.403	5.010

‡ For the difference between two entries

† Average of 37 $F_{6:8}$ lines

£ For the difference between F_2 and Mid-parent

Heritability for height had the largest range of the three traits measured; they ranged from 0.70 to 0.17. (Table 3.11) The correlations between height mid-parent heterosis and heritability, variance and GCV were all non significant.

Table 3.11 Genetic variance, Phenotypic variance, and Heritability of Height for each population.

Population	GCV	Genetic Variance	Phenotypic Variance	h^2	Standard Error of h^2
A	0.065	3.38	6.15	0.55	0.39
B	0.035	1.64	4.30	0.38	0.49
C	0.032	1.63	6.05	0.27	0.56
D	0.017	0.38	2.24	0.17	0.62
E	0.047	3.35	9.15	0.26	0.50
F	0.062	6.20	8.91	0.70	0.31

*Significant at significance level 0.05.

All three locations were planted at separate times. Because of the different planting dates, the maturity date was analyzed by location to see if there were any lines whose maturities would be to different from the rest of the population. No lines were found to be significantly outside the range of the two parents. (Appendix table 1, 2, 3) There was no significant variation for maturity found among $F_{6:8}$ lines.

Discussion

The yield data indicate that F_2 performance may be predictive of population performance. The only population of the six with a heterotic F_2 (population A) was also the only population that had inbred lines significantly higher yielding than their parents. In this case, if the breeder had screened his populations for F_2 heterosis he would have selected the population A as the most likely to give him transgressive segregation for yield. However, the sample size (37 inbred lines) is small. A larger sample in one of the other populations might have produced transegregating lines, but a larger sample in population A would likely have produced more as well. However the correlation between mid-parent yield heterosis and genetic variance of 0.592 was not significant. While this correlation is not statistically significant, it is noteworthy. The lack of significance may simply be because there was not enough populations to achieve significance. When yield mid-parent heterosis is plotted against genetic variance a general positive trend can be seen, but there are only six data points, therefore not enough to make a determination (Appendix 4).

There was no high-parent F_2 heterosis for seed size. Three of the six populations (A, B, C) showed mid-parent heterosis. All three of these populations also had inbred lines with seed sizes greater than the highest parent seed size (Table 3.6). In contrast to the yield data, a fourth population (E) which had numerical but not significant mid-parent for seed size did have lines that were significantly different from the high parent for seed size. There is no significant correlation between yield mid-parent heterosis and heritability or genetic variance or GCV. Even though population E had no significant heterosis,

selection upon F_2 heterosis would have lead to the selection of populations with lines significantly greater than the parents. If the purpose of selection upon F_2 heterosis is to increase the chances of finding an inbred line that is significantly greater than its parents, then these results support such a practice.

Of interest, population B showed very low genetic variance for both yield and seed size. Even with this low genetic variance, this population did have transgressive segregation for seed weight. Past research has shown that coefficient of parentage could only be used to predict crosses with a lower probability of generating high genetic variance (Gizlice et al. 1993; Cerna et al., 1997; Manjarrez-Sandoval et al., 1997). The genetic variance for yield, seed size, and height was regressed upon the Coefficient of Parentage (CP). The regression of CP upon yield and height were non-significant. However, the regression of CP upon the genetic variance for seed size has a significant positive relationship at p value 0.05.

Evaluation of F_2 performance within crosses of self-pollinated plants generally has been made on an individual plant basis. A comparison by Immer (1942) of the yield distribution of F_2 and parental single plants in three barley crosses revealed that the yield of an F_2 spaced plant was determined very largely by environmental factors. Excessive environmental effects when plants are space-planted lead to a low correlation that was generally found between the yield of single cereal plants in segregating populations and the mean of their progeny rows. It was concluded by Harrington (1932) that the actual yields of a given F_2 plant are not predictive of the heritable yielding ability of that plant.

Because of the difficulty in single plant evaluation, we regard F_2 bulk performance to be more reliable.

Weiss, et al (1947) tested the correlation between early generation selected soybean lines and pedigree selected lines and found the correlation to be low. Agronomic traits were measured in F_2 through F_5 bulk populations and compared to the means of F_5 lines surviving after pedigree selection. There was little agreement between the F_2 bulk populations and the F_5 lines (Weis, et al 1947). This experiment was conducted before the introduction of SSD so only selected advanced generation lines were compared to the F_2 bulk populations instead of a randomly developed inbred population. The current study tested both bulked F_2 populations and random sets of inbred lines derived from an SSD breeding program, thus giving a good estimate of genetic variation and a less biased comparison of F_2 bulk and inbred line performance.

Conclusion

Even if there is no heterosis found, or all genetic variance is additive, transgressive segregation can still occur (Matzinger, 1963). Bulked F_2 have been tested in other inbred crop species and have been found to be predictive in barley and wheat (Harrington, 1940; Immer, 1942). This study is evidence that, like the F_2 bulks in barley and wheat, F_2 bulks in soybeans can be predictive of SSD derived inbred line performance when in an intracross selection program. This study showed a positive of correlation between heterosis and population performance measures. It is important to note that transegregating lines would be expected in all the populations given the established error

rates. The fact that only one population had any and it was the population with heterosis is interesting. However there was no significant positive correlation between F_2 heterosis and genetic variance for any of the traits. Going forward more research should be done in order to test a larger number of populations. With a larger number of populations a more definitive determination could be made. Additionally the transegregating lines identified in this study should be retested to verify that they are indeed greater than their respective parents.

REFERENCES

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

Brim, C.A., and Elledge, C., 2001. Registration of 'Ransom' soybean. Crop Sci. 41:594a-595a.

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

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

Burton, J.W. and Brownie, C., 2006. Heterosis and Inbreeding Depression in Two Soybean Single Crosses. Crop Sci. 46:2643-2648.

Burton, J.W., Carter, Jr. T.E., and Huie., E.B. 1996. Registration of 'Holladay' soybean. Crop Sci. 36:467

Buss, G.R., Camper, Jr. H.M., Toane, C.W., 1988. Registration of 'Hutcheson' soybean. Crop Sci. 28:1024

Carter, T.E., Huei, E.B., Burton, J.W., Farmer, F.S., and Gizlice, Z., 1995. Registration of 'Pearl' soybean. Crop Sci. 35:1713.

Cerna, F.J., Cianzio, S.R., Rafalski, A., Tingey, S., Dyer, D., 1997. Relationship between seed yield heterosis and molecular marker heterozygosity in soybean. Theoretical and Applied Genetics. 95:460-467.

Gizlice, Z., Carter, T.E., Burton, J.W., 1993. Genetic diversity in North American soybean: II. Prediction of heterosis in F2 Populations of southern founding stock using genetic similarity measures. Crop Sci. 33:620-626.

Hallauer, A. R., Miranda, J.B., Quantitative Genetics in Corn breeding. Iowa State University Press, Ames IO. 1981.

Harrington, J.B. Predicting the value of a cross as indicated by bulk hybrid tests. Can. Jour. Res., 6:21-37. 1932

Harrington, J.B. Yielding capacity of wheat crosses as indicated by bulk hybrid tests. Can Jour Res., 18:578-584. 1940

Hartwig, E.E., and Epps, J.M., 1973. Registration of 'Forrest' soybean. Crop Sci. 13:287.

Hartwig, E.E., 1974. Registration of 'Tracy' soybean. Crop Sci. 14:777a.

Hartwig, E.E., and Epps, J.M., 1977. Registration of 'Centennial' soybean. Crop Sci. 17:979a.

Hartwig, E.E., and J.M. Epps. 1978. Registration of Bedford soybeans. Crop Sci. 18:915.

Hinson, K., and Hartwig, E.E., 1964. Registration of 'Bragg' soybean. Crop Sci. 4:664.

Immer, R.F. 1941. Relation between yielding ability and homozygosis in barley crosses. Jour. Amer. Soc. Agron., 33:200-206.

Immer, R.F. 1942. Distribution of yields of single plants of varieties and $_{F2}$ crosses of barley. Jour. Amer. Soc. Agron., 34:844-850.

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

Metzinger, D. F., 1963. Experimental estimate of genetic parameters and their applications in self-fertilizing plants. Symposium on Statistical Genetics and Plant Breeding, Raleigh, North Carolina. pp 253-276.

Nickell, C.D., Schwenk, F.W., and Schapaugh Jr., W.T., 1982. Registration of 'Douglas' soybean. Crop Sci. 22:160-a.

SAS Institute Inc., SAS 9.1.3 Help and Documentation, Cary, NC: SAS Institute Inc., 2000 2004.

Shipe, E.R., Mueller, J.D., Lewis, S.A., Williams, P.F., and Tomkins, J.P., 1997. Registration of 'Dillon' soybean. Crop Sci. 37:1983

Smith, T.J., and Camper, H.M., 1973. Registration of 'Essex' soybean. Crop Sci. 13:495-a.

Weiss, M.G., Weber, C.R., Kalton, R.R., 1947. Early generation testing in soybeans. Journal of the American Society of Agronomy. 39:791-811.

CHAPTER 3

Introduction

There are several examples of heterosis for yield in soybeans. In a review of 9 experiments comprising comparisons of 260 F_1s from biparental crosses with their parents in spaced plant field design, the mean percentage high parent heterosis of F_1 was 13.4% (Burton, 2008). Fifty-five percent of biparental cross combinations had F_1 yields greater than the high parent. The average percentage mid-parent heterosis of the F_1s was 25.7% for yield. Cerna et al. (1997) compaired 16 F_1s in a replicated spaced plant design (2 years, 2 locations) and found significant mid-parent and high-parent heterosis for 11 and 5 of the crosses for yield, respectively.

F_1 hybrids have also been tested in standard row plots used by breeders for estimating yield per unit area. Average high-parent heterosis was reported as high as 20% for the F_1 of 2 crosses (Brim and Cockerham, 1961). Nelson and Bernard (1984) reported an average 3.3% high- parent heterosis among 37 crosses. Burton and Brownie (2006) found 16% F_1 high-parent heterosis for yield in the biparental cross of Hutcheson x Holladay. F_2 bulks have also been tested in standard yield plots along with parents. In four such experiments, average mid- parent heterosis was found to be 8%, 11%, 9%, and 7% (Weiss et al., 1947; Loiselle et al., 1990; Gizlice et al. 1993; Manjarrez-Sandoval et al., 1997). Recently Friedrichs (2009) found of four F_1 hybrids tested, one had significant F_1 high parent heterosis of 30.7%.

An experiment attempting to explain the genetic basis for heterosis was conducted by Swanson-Wagner et al. (2006). Swanson-Wagner et al. (2006) used microarray technology to characterize the modes of gene action for 13,999 cDNAs in corn. Nearly 10% of the ESTs exhibited differential expression among the three genotypes. The majority of these, 78% (n=1,062), were statistically indistinguishable from additivity. Most of the remaining 22% (n=305) exhibited high-parent dominance. However, 2% (n=23) exhibited low-parent dominance, less than 1% (n=10) exhibited underdominance, and 3% (n=34) exhibited overdominance. These results were consistent with the hypothesis that multiple molecular mechanisms contribute to heterosis (Swanson-Wagner et al. 2006) however this experiment was done using F_1s so the result are possibly not representative actual genetic mechanisms. F2 segregation is needed in order to estimate genetic effects.

In soybeans, Schlueter et al (2004) found at least two genome duplication events using EST databases. Several previous studies had shown similar results with hybridization- based genetic maps (Shoemaker et al. 1996; Lee et al. 1999; Lee et al. 2001) and RFLP probes which detected extensive homoeologous relationships between linkage groups (Shoemaker et al. 1996). This genome duplication would contribute to the conservation of mildly deleterious genes (Husband and Schemske, 1996), which would partly explain heterosis in an F_1 hybrid that combined dominant favorable alleles at both loci, one contributed from each parent.

Genome regions have been studied in the past in order to identify regions that

affect yield. Herrmann et al. (2006) found two genome regions in red clover containing four or five QTLs for different seed yield components. Rafalski (2002) theorized that the elite germplasm of some crops may have been subjected to bottlenecks recently, increasing the amount of linkage disequilibrium (LD) present and facilitating the association of single nucleotide polymorphisms (SNPs) haplotypes at candidate gene loci with phenotypes. Whole-genome scans would help identify genome regions that are associated with interesting phenotypes if sufficient LD was present. Hyten et al. 2007 investigated the structure of LD in three regions of the genome varying in length from 336 to 574 kb. This analysis was conducted in four distinct groups of soybean germplasm: 26 accessions of the wild ancestor of soybean (Glycine soja Seib. et Zucc.); 52 Asian G. max Landraces, the immediate results of domestication from G. soja; 17 Asian Landrace introductions that became the ancestors of North American (N. Am.) cultivars, and 25 Elite Cultivars from N. Am. In G. soja, LD did not extend past 100 kb; however, in the three cultivated G. max groups, LD extended from 90 to 574 kb, likely due to the impacts of domestication and increased self-fertilization. The three genomic regions were highly variable relative to the extent of LD within the three cultivated soybean populations (Hyten et al. 2007). These experiments point toward blocks within the soybean genome potentially important in the heterosis that has been identified in previous hybrid studies. If markers could be identified in a heterotic population then the regions around those markers that are in LD could be further investigated.

The objective of this experiment was to identify regions of the soybean genome that are responsible for the high parent heterosis for yield of the Hutcheson x Holladay

hybrid. $F_{2:3}$ families from the cross were yield tested in field experiments. A molecular marker on each arm of every chromosome pair were used with phenotypic data to identify regions associated with yield. A second year of phenotype data was collected on the following generation in order to tentatively confirm the significance of markers identified in the initial evaluation.

Materials and Methods

F_1 seed from the Hutcheson x Holladay cross was used to generate F_2 seed in the winter nursery in Puerto Rico in 2007. This F_2 seed was used for tissue sampling for DNA analysis and to generate the $F_{2:3}$ lines for yield trialing.

As a way of sampling the entire genome, a single polymorphic marker was identified on each arm of the 20 chromosome pairs. Marker identification was done by doing DNA screens for selected SSR markers. SSR markers were used because of the availability of genetic maps with SSRs and the cooperation of supporting laboratories. Starting at the end of each chromosome and working towards the center, 48 markers were selected at a time and put into a single 96 well PCR plate. The first row of the plate contained one parent, while the second row contained the other parent. This plate layout resulted in the two parents being paired next to each other when placed into the agarose gel. This made identification of polymorphic markers a matter of looking for paired marker bands that were off set. 40 SSR markers were selected to screen the F_2 individuals (2 per 20 linkage groups) (Table 4.1).

Table 4.1 The selected SSR markers that are polymorphic in the Hutcheson x Holladay cross, their linkage group and centimorgan position from the integrated genetic linkage map of soybean (Song et al., 2004).

Marker	LG	cM Pos	Marker	LG	cM Pos	Marker	LG	cM Pos	Marker	LG	cM Pos
Satt684	A1	3.54	Satt357	C2	151.91	Sat_168	G	3.9	Satt513	L	106.37
Satt200	A1	92.89	Sat_201	D1a	56.08	Satt394	G	43.38	Satt220	M	56.29
Satt424	A2	60.59	Satt436	D1a	70.69	Satt635	H	4.88	Satt336	M	133.83
Satt158	A2	115.25	Sat_211	D1b	38.04	Satt181	H	91.12	Satt530	N	32.85
Satt597	B1	73.77	Sat_192	D1b	135.26	Satt292	I	82.78	Sat_304	N	77.1
Sat_331	B1	125.74	Satt154	D2	57.07	Satt440	I	112.7	Satt420	O	49.71
Sat_264	B2	12.56	Satt256	D2	124.31	Satt285	J	25.51	Satt581	O	106.03
Satt066	B2	78.84	Satt212	E	32.27	Satt547	J	67.79			
Satt646	C1	70.52	Satt185	E	44.76	Satt349	K	42.39			
Satt338	C1	123.79	Satt325	F	2.23	Satt260	K	80.12			
Satt281	C2	40.3	Sat_313	F	91.87	Satt143	L	30.19			

In the winter of 2005, F_2 seed was generated from the F_1 seed of the heterotic cross of 'Hutcheson' (Buss et al., 1988) and 'Holladay' (Burton et al., 1996). The DNA needed for genotyping the F_2s was isolated from leaf tissue samples. The leaf tissue samples were collected from F_2 plants grown in the summer of 2006 in Clayton, North Carolina and the following winter in Puerto Rico. To ensure there would be enough leaf tissue sampled F_2 plants that would produce a viable amount of seed, a second set of F_2s were grown in the winter nursery. Enough viable seed was needed for multiple location yield testing in common bordered rows the following year. Of the plants from Clayton, 312 had enough seed to produce $F_{2:3}$ lines. 150 F_2 plants were grown in Puerto Rico in the winter of 2006 to supplement the 312 plants harvested from Clayton. Tissue samples were collected from newly expanding leaves. Upon collection of the tissue, using latex gloves to protect against contamination, each sample was placed in a single 150 ml centrifuge tube and labeled to match the plant from which it was collected. The sample containing centrifuge tubes were placed directly into Styrofoam containers of liquid nitrogen, instantly freezing the sample. A very low temperature was maintained in order to preserve it. The frozen samples were then taken back to the lab and placed in a -80^o C freezer.

DNA isolation of the F_2 leaf tissue samples was done using 96 well blocks. A grid was made up to show the location of each of the samples. A part of the tissue of each sample was placed in each of the 96 wells for DNA extraction. Before any tissue was placed in the blocks, a single sterile BB was placed in each well to assist in the grinding of the frozen sample. When placing tissue into each of the blocks, 10-15 centrifuge tubes were removed from the freezer at a time and placed in a Styrofoam container of liquid

nitrogen. The centrifuge tubes were then opened, and then a part of the tissue sample removed and placed in its appropriate place in the 96 well block. The tube was closed and returned to the freezer. This process was repeated until all the genotypes were placed into their respective blocks. In each of the blocks, the two parents of this F_2 population were entered twice. The two parents were placed such that during the marker screening each parent would be seen in either the top or bottom half of a 96 well PCR reaction. In total four 96 well blocks were prepared for DNA extraction.

The four 96 well blocks of plant tissue were then put through a modified CTAB (hexadecyltrimethylammonium acid) DNA extraction protocol (Keim et al., 1988) and resuspended in TE buffer (buffer solution of Tris and ethylenediaminetetraacetic acid (EDTA)). 50 microliters of TE was added to each well of a 96 well block, lids were replaced onto each well, and the boxes were placed into the refrigerator until the following day to allow the pellets to be re-suspended. All four labeled boxes were then placed into a -80^{o} C freezer for longer storage.

This extraction was repeated a second time half way through the population screening in order to obtain more DNA, yielding an average of 50 nanograms DNA per microliter. To save the integrity of the DNA small portions, 10 microliters at a time, were removed from each box and placed into labeled dilution boxes. The DNA solution was diluted 10:1 to yield 5 nanograms per microliter dilution.

For genotyping the population, 384 and 96 well PCR plates were used. When 384 plates were used the entire population could be genotyped at the same time for one

marker. With the use of the 96 well plates, 4 plates were run for each marker. For each of the PCR plates, 5 microliters of DNA was added and then 10 microliters of master mix was added. The master mix for the 384 well PCR plates included: 2064 microliters of DNA grade H_2O, 600 microliters of 10X buffer, 800 microliters of 15 mM $MgCL_2$, 360 microliters of 3.12 mM dNTPs (Deoxyribonucleotide triphosphates), 100 microliters of Taq polymerase, and 18 microliters of the primer. The master mix for the 96 well PCR plates was include: 516 microliters of DNA grade H_2O, 150 microliters of 10X buffer, 200 microliters of 15 mM $MgCL_2$, 90 microliters of 3.12 mM dNTPs, 25 microliters of Taq polymerase, and 4.5 microliters of the primer. For each of these master mixes the amount was slightly larger than what was required to fill all of the wells. The larger amount allowed multichannel pipetting of the master mix into the PCR plates using a 12 channel pipette.

To make the 500 milliliters of 10x buffer, 100 milliliters of 1 Molar Tris with pH 8.4 was added with 125 milliliters of 2 Molar KCL and 275 milliliters of ddH_20 (double distilled water) and stored in the refrigerator at 4 C. The 15 mM $MgCl_2$ was made using 1.5 milliliters of 1 Molar $MgCl_2$ and 98.5 milliliters of ddH_2O. The dNTPs arrived in separate tubes of 100 mM each. All four tubes would be combined to form a 25 mM solution of dNTPs. 63 microliters of 25 mM dNTP mix would be added to 437 microliters of DNA grade H_2O to yield 500 microliters of the required 3.15 mM dNTP mix.

After each of the plates (either 384 or 96 well) had been prepared, they were put into a thermocycler. A customized thermocycle schedule was used (Appendix table 1).

Once each PCR plate had finished it's cycling time it was placed into a refrigerator to be stored at 4^{o} C until it migrated on a gel for scoring.

A 4 % Super Fine Resolution (SFR) agarose gel was used for scoring the DNA bands. To make a single batch of gel, material that can be used in four 96 well gels, 32 grams of SFR agarose was dissolved in 800 milliliters of 1X TBE in a 1 liter Erlenmeyer flask. The dissolving was achieved by heating the mixture in the microwave until the liquid is clears. Once all the SFR has dissolved, 5 microliters of ethidium bromide is added per 100 milliliters of solution or 40 microliters. After each time the gel was used, 20 microliters of ethidium bromide were added back to replenish what had been lost.

The solidified gel mixture was heated in a microwave until completely liquefied. The heated liquid agarose was then poured into prepared trays. After the well combs had been added, they were allowed to cool for approximately one hour or until completely solid. Once the gel's had solidified, they were loaded into gel boxes filled with 1X TBE (Tris-Borate- EDTA).

96 or 384 well plates that had finished PCR were loaded with a loading dye. 4 microliters of 6X loading dye was added to each of the wells. The plates were centrifuged to settle all the liquid into the bottom of the wells. The PCR product and loading dye were then multichannel pipetted into the gels in the gel boxes. The gel boxes were set at 100 volts for 3 hours. The gel boxes were checked after 2 hours; if the bands had migrated enough to score, a picture was taken and the gel was broken and placed back in its Erlenmeyer flask. If the bands had not migrated enough the gel was placed back into the

gel box and checked an hour later. This was repeated until the SSR bands had migrated enough to score. A digital picture was taken of the gel under ultraviolet light and scored. Scoring was done by identifying homozygous and heterozygous individuals. The homozygous individuals were further classified as to parent of origin. 360 F_2 individuals were scored for all 40 polymorphic markers.

Field Experiment

360 F_2 plants of the Hutcheson x Holladay cross, each of which had at least 320 seed, enough for testing a single replication at 2 locations of yield testing, from the summer and winter of 2006 were selected for a 400 plot incomplete block yield trial. Three row plots were used. In order to allow for enough seed in three locations, a common border of a 50:50 mixture of the two parents was used. There were 20 incomplete blocks, each included 18 $F_{2:3}$ lines, and the two parents. The 18 unique genotypes in an incomplete block were consistent over locations. The field location of each incomplete block and genotype within each block were randomized within each replication and location. The plots were planted using a three row research planter. The rows were 0.97 meters wide and 5.8 meters long and were end trimmed before harvest to establish a uniform row length. At mid-pod filling stage, the plots were trimmed to a length of 4.9 meters in Clinton and Plymouth and a length of 4.3 meters in Caswell. The middle row of each plot was harvested at maturity.

In the summer of 2007, a single replication of this experiment was grown at two locations: Kinston and Plymouth NC. Due to dry weather conditions the data from the Clinton location was unreliable and was discarded. Flowering date, date of maturity,

height, and lodging were noted for each plot at each location. After harvest, total seed weight (converted to kilograms per hectare) and grams per 100 seed were determined for each plot.

The seed from this test were combined by genotype and sampled to form $F_{2:4}$ lines. In 2008, the experiment was repeated to test any significant dominant markers identified in the $F_{2:3}$ lines trials. These 360 $F_{2:4}$ lines were yield tested in the summer of 2008 at two locations (Kinston, and Plymouth) using the same field experiment design and row dimensions, however parents were not used as borders because sufficient seed was available for a full three row plot. The same lines within each block from 2007 were maintained in this trial. Within these two locations, an incomplete block design with two replications was used. After harvest, total seed weight (converted to kilograms per hectare) and seed sizes (grams per 100 seeds) were determined for each plot.

Marker Analysis

The 40 polymorphic markers were used to screen the population and develop a genotype for each F_2 individual. The screening was done using both 384 and 96 well plates. The screening was started with 384 well plates, but due to a very high failure rate of PCR in these 384, the protocol was switched to using only 96 well plates. Digital pictures were taken of each agarose gel and scored; the Hutcheson allele was scored as 0, the Holladay allele was scored as 2, and the heterozygote was scored as 1. The data were entered into one Excel spreadsheet which was used for later data analysis when combined with the phenotypic data.

Data Analysis

The phenotypic data were combined over locations within years and analyzed using PROC MIXED in SAS (SAS Institute, 2004). Location, block, and location by block interaction, were treated as random while code within block was treated as fixed. Means, for analysis with marker data, of the $F_{2:3}$ and $F_{2:4}$ lines were estimated using the LSMEANS statement. When analyzing for additive effects, the yield data were combined over years because the expected segregation ratios were identical. For the dominance tests, the yield data were kept separate by year because of the differing expected segregation ratios of the $F_{2:4}$ lines derived from heterozygous $F_{2:3}$ lines.

This marker data was paired with the least square means obtained from the 2007 and 2008 yield data. Single marker analysis was done using PROC GLM in SAS for all 40 markers on yield and seed weight. Additionally F2Epistacy.sas was used to test for epistatic interactions between loci (Holland, 1998). The single marker analysis was done at significance level 0.05 while the epistatic interactions were done with p value 0.001 because of a Bonferroni correction for multiple tests.

Results

Yield

Analysis of variance for yield showed significant variance for yield in the $F_{2:3}$ lines in 2007. When comparing the line(block) means square for the $F_{2:3}$ and $F_{2:4}$ there is a slight decrease, with this decrease the mean square for line(block) is no longer significant at significance level 0.05.

Table 4.2 Yield (kg ha^{-1}) Means Squares for $F_{2:3}$ and $F_{2:4}$ lines tested at two NC locations in 2007 and 2008.

Source	Df	$F_{2:3}$	Df	$F_{2:4}$
Location	1	96659336*	1	181955169
Block	19	1543691	19	4308176*
Line(Block)	339	192763*	338	1908465
Location x Block	19	1598304*	19	1200072
Rep(Location)	.	.	2	3074223
Block x Rep(Location)	.	.	38	2795312
Location x Line(Block)	336	147748	338	2083548
Residual	.	.	663	2065668

*Significant at significance level 0.05

Satt292, Satt530, and Satt420 tested significant for additive effects. The positive allele at locus Satt292, Satt530, and Satt420 came from Hutcheson with effects of 32.38 kg ha^{-1}, 32.65 kg ha^{-1}, and 35.73 kg ha^{-1} respectively. Satt336 is not significant at p value

0.05 but is significant at p value 0.06, with an effect of 26.89 kg ha^{-1} from Hutcheson. (Table 4.3)

Table 4.3 Significant Additive estimates in kilograms per hectare for F_2 derived lines from the cross of Hutcheson and Holladay combined over two years and two locations for two SSR markers per linkage group (Kinston and Plymouth)‡.

Marker	Linkage Group	Arm	Additive Estimate	Standard Error
Satt292	I	Upper	69	30
Satt336	M	Lower	57£	29
Satt530	N	Upper	69	29
Satt420	O	Upper	76	38

£Significant at significance level 0.06

‡One replication per location in 2007, two replications per location in 2008

Four markers had significant positive dominance effects in 2007. The positive dominance effect observed in 2007 for Satt154, Satt181, Satt530, and Sat_304 came from Hutcheson. Sat_304 had significant overdominance in 2007. Three different markers, Satt684 and Satt143 with a positive dominance effect, and Satt158 with a negative dominance effect were observed in 2008. Satt684 and Satt143 came from Hutcheson, with Satt158 coming from Holladay. None of the dominant markers were significant in both years. When comparing the additive value estimated for Satt530 with the two dominance estimates the marker appears to have overdominance. However these are estimates cannot be compared to each other because the additive estimate is derived from

data combined over both years. (Table 4.4)

Table 4.4 Significant Dominance estimates adjusted for inbreeding in kilograms per hectare for F_2 derived lines from the cross of Hutcheson and Holladay for 2007 and 2008, combined over two locations for two SSR markers per linkage group (Kinston and Plymouth)‡.

Marker	Linkage Group	Arm	$F_{2:3}$ Estimate	$F_{2:3}$ Standard Error	$F_{2:4}$ Estimate	$F_{2:4}$ Standard Error
Satt684	A1	Upper	315	163	1316*	651
Satt158	A2	Lower	187	165	-1365*	659
Satt154	D2	Upper	324*	163	-900	722
Satt181	H	Lower	451*	181	705	758
Satt143	L	Upper	-177	176	1611*	734
Satt530	N	Upper	319*	162	1076	643
Sat_304	N	Lower	735*£	194	-812	700

£Significant overdominance at significance level 0.05

*Significant at significance level 0.05

The SSR marker, Sat_304 from the lower arm of the N linkage group, was tested

for overdominance in the 2008 yield data. An estimate of 59.94 kg ha^{-1} was found to be significant for overdominance with a p value of 0.02.

After the single marker analysis was done, the $F_{2:3}$ yield data were tested for all epistatic interactions. Table 4.5 lists all interactions that were significant at p value 0.005. With a Bonferroni adjustment for the 780 tests, a p value of 0.001 should be used. With the Bonferroni adjustment, the interaction between Sat_331 and Satt185 was a significant dominance by additive interaction and the interaction of Satt260 by Satt420 was a significant additive by additive interaction. (Table 4.4)

Table 4.5 Significant two locus epistatic interactions in kilograms per hectare for F_2 derived lines from the cross of Hutcheson and Holladay for 2007, combined over two locations for two SSR markers per linkage group (Kinston and Plymouth)‡.

Locus 1	Locus 2	F Value for AxA	P > F AxA	F Value for DxA	P > F DxA
Sat_313	Satt424	5.8564	0.016	0.2571	0.613
Sat_331	Satt185	0.2181	0.641	18.6533	0.000†
Satt143	Satt200	1.9157	0.167	8.7278	0.003
Satt260	Satt420	16.1911	0.000†	0.0167	0.897
Satt285	Satt357	1.409	0.236	9.0943	0.003
Satt292	Satt338	1.836	0.177	9.9489	0.002

†Significant at p = 0.001

‡One replication per location in 2007.

The epistasis analysis for the $F_{2:4}$ 2008 data yielded only one additive by additive

interaction of Satt635 and Satt646 with a p value of 0.002. This p value is not significant with a Bonferroni correction p value of 0.001.

Seed Size

There was no significant variance in the population in 2008 for seed size. (Table 4.6) No significant SSR markers were found to be associated with seed size for either additive or dominance effects.

Table 4.6 Seed size (grams per 100 seed) Means Squares for $F_{2:4}$ lines tested at two NC locations in 2008.

Source	Df	$F_{2:4}$
Location	1	431
Block	19	1365
Line(Block)	338	1691
Location x Block	19	1463
Rep(Location)	2	2360
Block x Rep(Location)	38	1722
Location x Line(Block)	338	1682
Residual	664	1698

*Significant at significance level 0.05

Discussion

The additive marker effects should not have been affected by the changes in segregation within a single F_2 bulk. If an F_2 line were homozygous for this allele, then all of its progeny would remain homozygous for that allele, which therefore means there would be no change in effects due to a change in the segregation of alleles as a result of further inbreeding. The contrast developed for testing additive effects tests the two homozygous classes against each other. The additive effects, tested over years, did produce three significant additive quantitative trait loci (QTL): Satt292, Satt530, and Satt420. All three QTL had a high parent value derived from Hutcheson. None of the three SSR markers have been previously reported as yield QTL in Soybase and are therefore unconfirmed yield QTL. (http://soybase.org; verified May 5, 2009)

The 2007 and 2008 data both show genetic evidence of dominance. The dominant markers were inconsistent across years. The dominance loci could change in value with inbreeding. The F_2 segregates ¼ AA ½ Aa ¼ aa, the F_3 segregates 3/8 AA ¼ Aa 3/8 aa, and the F_4 segregates 7/16 AA 1/8 Aa 7/16 aa. With these segregation ratios, the $F_{2:3}$ derived from a heterozygous F_2, would have two phenotypic classes; ¾ A_ and ¼ aa, this would change in the $F_{2:4}$ to 5/8 A_ and 3/8 aa. The change in segregation ratios would mean that there should be a steady decline in yield for the heterozygous F_2s when the dominance value is positive. This decline could account for the loss of significance for several of the weaker dominance markers (Satt154 and Satt530 both with p value of ~0.04). However, the decline does not account for the appearance of the dominant

markers in the $F_{2:4}$ that were not seen in the previous generation (Satt684, Satt158, and Satt143).

The overdominant loci, Sat_304, would have declined in yield with the additional year of inbreeding from the $F_{2:3}$ to the $F_{2:4}$. With inbreeding, the number of heterozygotes within that single $F_{2:4}$ field plots would be half that in the $F_{2:3}$ field plot. The change in the number of heterozygotes would greatly decrease the impact of overdominance on the overall plot yield. In the 2008 $F_{2:4}$ yield data the dominance value is no longer significant. In fact it switched from a positive value to a nonsignificant negative value (-45.17).

There was inconsistency between years for marker results. The inconsistency may have been due to an inability to get good yield estimates from $F_{2:3}$ lines due to lack of seeds for testing. Also, whenever a common border is used, inter-genotypic competition between the borders and center row can bias the yield estimates. If QTL could be confirmed in the next generation, after a meiotic cycle, the QTL would be more efficient in a marker assisted selection program (Asghari et. al 2008). There are more than 900 QTL reported in SoyBase for the various quantitative traits, not all of these QTL have been confirmed, and the results of those that have been studied has been inconsistent. For example, Diers et al. (1992) reported that none of the identified QTL conditioning iron deficiency chlorosis in soybean was effective for divergent selection among lines from the same cross that were not used in the original QTL mapping study. Finally the inconsistency may have been due to significant genotype by year interactions.

Not only has there been some difficulty in confirming QTL in different genetic

backgrounds, but confirming QTL across environments can be difficult. Brummer et al. (1997) identified QTL for soybean seed protein and oil content using eight distinct populations. They reported that the phenotypic effect of some QTL was sensitive to the environment in which they were evaluated, but did detect environmentally stable QTL. In maize (*Zea mays* L.), results from three independent experiments repeated in the same genetic background revealed that the QTL identified were not consistent (Beavis et al., 1994; Beavis, 1994). Beavis (1994) suggests that the factors confounding analysis could be population structure, sources of parental lines, different sets of environments, and sampling of progeny. With this experiment, the different sets of environments, the common border used in the 2007 trial, and the sampling of the $F_{2:3}$ lines to generate $F_{2:4}$ lines could be the confounding factors.

Fasoula et al. (2004) did a study to confirm previously reported QTL for seed protein, seed oil, and seed weight by using two independent populations, PI97100 x 'Coker 237' and Young x PI416937, with SSR and RFLP markers. Single-factor analysis of variance was used to verify the QTL that had significant ($P \leq 0.01$) associations. In the PI97100 x 'Coker 237' population, two (*cqProt-001* and *cqProt-002*) of four previously described QTL for seed protein, two (*cqOil-001* and *cqOil-002*) of three QTL for oil content, and none of three QTL for seed weight were confirmed in the independent population. In the Young x PI416937 population, none of the three previously reported QTL for protein was confirmed. One (*cqOil-003*) of three QTL for oil content and two (*cqSd wt-001* and *cqSd wt-002*) of three QTL for seed weight were verified. This study emphasized the importance of validating previously reported QTLs because the

unconfirmed QTL may have been false positive or they may have been specific for the sample of lines used in the original populations.

Conclusions

The current experiment illustrates many of the difficulties that have been faced in the detection and confirmation of QTL in a crop. First, as stated by Beavis (1994), differing environments and sampling errors when generating the $F_{2:4}$ lines could lead to false positive QTLs or failure to confirm significant QTLs from the $F_{2:3}$ lines. A combined year analysis of the $F_{2:3}$ and $F_{2:4}$ lines were not considered valid because of the segregation as a result of a meiotic event. This segregation is important, because dominance has been proposed in soybeans by Burton and Brownie (2006). Without the possibility of dominance, the differing segregation ratios would be of no significance. Second, Fasoula et al. (2004) points out that QTL detected in only one experiment, without confirmation, may simply be false positives or unique to that specific genotype and environment. Finally, the small sample of markers used in this experiment covered a very small portion of the genome.

The significant dominant markers found in each year would not be considered significant QTL because they were not confirmed in both years. If there had been a significant marker it was probably covered by the confounding effect of genotype by year interaction. This study is important in that it looks at a population specifically because of its heterotic nature. A larger sample marker sample size may be needed to adequately

evaluate the genome regions associated with yield. The additive markers need to be confirmed in a second population that could be replicated and tested over several locations. The markers could also be confirmed by developing NILs from this population and determining if the markers are validated.

REFERENCES

Asghari, A., Mohammadi, S.A., Mohaddam, M., and Shokuhian, A.A., 2008 Identification of SSR and RAPD markers associated with QTLs of winter survival and related traits in Brassica napus L." African Journal of Biotechnology 7.7: 897-903.

Beavis, W.D., 1994. The power and deceit of QTL experiments: Lessons from comparative QTL studies. p. 250–266. *In* Proc. of the 49th Annual Corn and Sorghum Industry Research Conference. ASTA, Washington, DC.

Beavis, W.D., O.S. Smith, D. Grant, and R. Fincher. 1994. Identification of quantitative trait loci using a small sample of top crossed and F4 progeny of maize. Crop Sci. 34:882–896.

Brim, C.A., and Cockerham, C.C., 1961. The inheritance of quantitative characters in soybeans. Crop Sci. 1:187-190

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

Burton, J.W., Carter, Jr. T.E., and Huie., E.B. 1996. Registration of 'Holladay' soybean. Crop Sci. 36:467

Burton, J.W. and Brownie, C., 2006. Heterosis and Inbreeding Depression in Two Soybean Single Crosses. Crop Sci. 46:2643-2648.

Burton, J.W., 2008. A review of heterosis in Soybean. Unpublished manuscript.

Buss, G.R., Camper, Jr. H.M., Toane, C.W., 1988. Registration of 'Hutcheson' soybean. Crop Sci. 28:1024

Cerna, F.J., Cianzio, S.R., Rafalski, A., Tingey, S., Dyer, D., 1997. Relationship between seed yield heterosis and molecular marker heterozygosity in soybean. Theoretical and Applied Genetics. 95:460-467.

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

Fasoula, V.A., Harris, D.K., and Boerma R.H., 2004 Validation and designation of quantitative trait loci for seed protein, seed oil, and seed weight from two soybean populations. Crop Sci 44:1218-1225.

Friedrichs, M.F. 2009. Heterosis and Inbreeding depression in *Glycine max*. Unpublished manuscript.

Gizlice, Z., Carter, T.E., Burton, J.W., 1993. Genetic diversity in North American

soybean: II. Prediction of heterosis in F2 Populations of southern founding stock using genetic similarity measures. Crop Sci. 33:620-626.

Herrmann, D., Boller, B., Studer, B., Widmer, F., and Kolliker, R., 2006. QTL analysis of seed yield components in red clover (Trifolium pratense L.). TAG Theoretical and Applied Genetics 112.3:536-545.

Holland, J. Epistasy: A SAS Program for detecting Two-Locus Epistatic Interactions Using Genetic Marker Information. Journal of Heredity 1998, Vol 89:374-375

Husband, B.C., Schemske, D.W. 1996: Evolution of the magnitude and timing of inbreeding depression in plants. Evolution, 50, 54-70.

Hyten, D. L., Choi, I., Song, Q., Shoemaker, R.C., Nelson, R.L., Costa, J.M., Specht, J.E., and Cregan, P.B., 2007. Highly variable patterns of linkage disequilibrium in multiple soybean populations. Genetics 175.4: 1937-1944.

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

Lee, J.M., Bush, A., Specht, J.E., and Shoemaker, R. 1999. Mapping duplicate genes in soybean. Genome, 42: 829-836

Lee, J.M., Grant, D., Ballejos, C.E., and Shoemaker, R. 2001. Genome organization in dicots. II. Arabidopsis as a 'bridging species' to resolve genome evolution events among legumes. Theor. Appl. Genet. 103: 765-773

Loiselle, F., Voldeng, H.D., Turcotte, P., St. Pierre, C.A. 1990. Analysis of agronomic characters for an eleven parent diallel of early-maturing soybean genotypes in eastern Canada. Canadian Journal of Plant Science, 70, 107-115.

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

Nelson, R. L. and Bernard, R. L. 1984, Production and performance of hybrid soybeans. Crop Science, 24:549-553.

Rafalski, A. 2002. Applications of single nucleotide polymorphisms in crop genetics. Current Opinion in Plant Biology 5.2: 94-100.

SAS Institute Inc., SAS 9.1.3 Help and Documentation, Cary, NC: SAS Institute Inc., 2000-2004.

Schlueter, J.A., Dixon, P., Granger, C., Grant, D., Doyl, J.J., Shoemaker, R.C. 2004. Mining EST databases to resolve evolutionary events in major crop species. Genome, 47, 868-876.

Shoemaker, R., Polzon, K., Labate, J., Specht, J., Brummer, E.C., Olson, T., Young, N.,

Concibido, v., Wilcox, J., Tamulonis, J., Kochert, G., and Boerma, H.R., 1996. Genome duplication in soybean (*Glycine* subgenus *soja*). Genetics, 144:329-338.

Song Q.J., Marke, L.F., Shoemaker, R.C., Lark, K.G., Concibido, V.C., Delannay, X., Specht, J.E., and Cregan, P.B., 2004 A new integrated genetic linkage map of the soybean. Theor Appl Genet 109:122-128.

Swanson-Wagner R.A., Jia, Y., DeCook, R., Borsuk, L.A., Nettleton, D., Schnable, P.S., 2006. All possible modes of gene action are observed in a global comparison of gene expression in a maize F1 hybrid and its inbred parents. PNAS, 103, 6805-6810.

Weiss, M.G., Weber, C.R., Kalton, R.R., 1947. Early generation testing in soybeans. Journal of the American Society of Agronomy. 39:791-811.

APPENDICES

Appendix 1 ANOVA for all crosses across locations, showing significant genotype by location interaction.†

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Genotype	24	2503958	104332	1.47	0.1821
Location	2	3290380	1645190	11.93	0.0091
Rep(Location)	3	489795	163265	4.12	0.0096
Genotype*Location	22	1570843	71402	1.8	0.0339
Residual	68	2693207	39606	.	.

†There is no year effect because no location was repeated over years.

Appendix 2 ANOVA for all crosses in 2007.

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Genotype	24	1789847	74577	1.77	0.0468
Rep	2	60954	30477	0.72	0.4912
Residual	48	2027666	42243	.	.

Appendix 3 ANOVA for all crosses across location in 2008.

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Genotype	24	3115008	129792	2.88	0.0060
Location	1	306456	306456	0.87	0.5281
Rep(Location)	1	360570	360570	6.82	0.0153
Genotype*Location	24	1080854	45036	0.85	0.6516
Residual	24	1269372	53890	.	.

Appendix 4 Maturity date, Lodging and Height means for all Genotypes across locations.

Genotype	MD mean†	LOD mean‡	HT Mean§
G98-1053	23.50	2.50	34.33
N99-244	29.00	2.00	32.00
G98-1053 x N99-244 F_1	24.67	2.33	37.67
G98-1053 x N99-244 F_2	26.00	2.17	36.00
G98-1053 x N99-244 F_3	25.83	2.33	40.00
G98-1053 x N99-244 F_4	26.00	2.33	45.33
5601T	13.00	2.17	33.33
N00-370	22.83	2.67	36.00
5601T x N00-370 F_1	19.50	2.17	36.00
5601T x N00-370 F_2	22.50	2.50	37.33
5601T x N00-370 F_3	20.00	2.17	38.67
5601T x N00-370 F_4	23.83	2.50	38.00
Hutcheson	10.83	2.67	33.33
Hutcheson x Holladay F_1	12.17	2.00	27.33
Hutcheson x Holladay F_2	11.67	1.83	32.33
Hutcheson x Holladay F_3	11.33	2.00	30.33
Hutcheson x Holladay F_4	13.50	2.17	30.33
Holladay	10.17	2.00	29.00
Satelite	23.50	2.33	42.00
Satelite x Holladay F_1	21.50	2.33	36.67
Satelite x Holladay F_2	23.33	2.67	41.67
Satelite x Holladay F_3	21.67	2.83	41.33
Satelite x Holladay F_4	22.83	2.50	43.33

†Maturity Date mean in days after July 1st.

‡Lodging score mean with a scale of 1-4

§ Height mean in centimeters.

Appendix 5 Maturity date means for Clayton by population in 2008

	A	B	C	D	E	F
♀ Parent	21	25.5	30	30	25.5	30
♂ Parent	30	25.5	21	30	30	30
F_2	25.5	29.9	30	30	30	30
Mid-Parent	25.5	25.5	25.5	30	27.75	30
LSD .05‡	6.0009	6.0009	6.0009	6.0009	6.0009	6.0009

‡ For the difference between two entries

Appendix 6 Maturity date means for Clinton by population in 2008

	A	B	C	D	E	F
♀ Parent	7	17	23	27	23	23
♂ Parent	17	7	7	17	7	17
F_2	17	17	17	23	17	17
Mid-Parent	12	12	15	22	15	20
LSD .05‡	6.1241	6.1241	6.1241	6.1241	6.1241	6.1241

‡ For the difference between two entries

Appendix 7 Maturity date means for Plymouth by population in 2008

	A	B	C	D	E	F
♀ Parent	13.5	23	29	25.5	25.5	28
♂ Parent	23	15.5	14.5	22.5	18	23
F_2	20.5	18	23	25.5	25.5	28
Lines†	16.5	19.25	21.5	24	21.75	25.5
LSD .05‡	5.3152	5.3152	5.3152	5.3152	5.3152	5.3152

‡ For the difference between two entries

Appendix 8 Thermocycle program used for Polymerase chain reaction.

Temperature	Time
1. 95C	2 minutes
2. 92C	30 sec
3. 49C	30 sec
4. 68C	45 sec
5. 39X return to step 2	
6. 72C	5 minutes
7. 4C	30 minutes
8. 10C	10 minutes
9. END	