ABSTRACT

RUFF, LEAH ANNE. Heterosis as Evidence of Yield Alleles from Wild Soybean. (Under the direction of Dr. Thomas E. Carter, Jr.).

To guarantee future global food security and sustainable crop production, there is strong need for broadening the relatively narrow genetic base of the soybean (*Glycine max* [L. Merr.]) germplasm and looking for new resources to develop soybean cultivars. Wild soybean (*Glycine soja* [Sieb. and Zucc.]) may be an excellent source of new agronomic alleles for this purpose. Heterosis for yield is proposed as a signal for detecting yield alleles in wild soybean. The objectives of this study were to 1) examine heterosis in F_2 bulks derived from the cross of parental cultivar N7103 and 19 breeding lines, which were developed from a cross between *G. max* cultivar N7103 and *G. soja* accession PI 366122, 2) assess whether heterosis could detect the presence of yield alleles derived from wild soybean, and 3) ascertain the possibility of predicting heterosis using genetic markers and agronomic traits. The individual F_2 bulked seed were yield tested in replicated bordered row plots at four locations in 2012. Two cross combinations showed significant ($P \leq 0.05$) percent midparent heterosis for yield (+9% and +10%). One cross showed significant heterosis over a theoretical maximum that could be expected based on dominance effects coming from only the *G. max* parent. Thus, these results suggest that yield alleles reside in wild soybean. The percent of *G. soja* alleles present in each breeding line, based on 558 single nucleotide polymorphic markers, positively and significantly ($P \leq 0.05$) associated ($r = 0.60$) with midparent heterosis. Genetic distance may be used as a guide for plant breeders when incorporating diverse germplasm into breeding programs and help predict future agronomic performance.

Heterosis as Evidence of Yield Alleles from Wild Soybean

by
Leah Anne Ruff

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

Crop Science

Raleigh, North Carolina

2013

APPROVED BY:

_____	_____
Dr. Thomas E. Carter, Jr. Committee Chair	Dr. Lilian Miranda
_____	_____
Dr. David Dickey	Dr. Julian Chaky

DEDICATION

To my family.

This thesis is dedicated especially to the family of farmers before me, my father Edmund Ruff III and my grandfathers Edmund Ruff II and Marvin Schulte. And to their wives, my mother Barbara Ruff and my grandmothers Estelle Ruff and Lois Schulte. Without the limitless support and love of my family and their passion for agriculture, which was instilled into me, this thesis would have never been possible.

BIOGRAPHY

Leah Anne Ruff was born December 14, 1988. She is the daughter of Edmund and Barbara Ruff. She grew up in the small town of Farmersburg, Iowa and within four miles of her grandmother's farm. The family farm was where Leah learned about and experienced agriculture. While growing up, she exhibited livestock, crop production, and horticulture projects at the Clayton County fair and Iowa State fair through the 4-H and FFA programs. She attended MFL MarMac Community Schools in Monona, IA and graduated in 2007. She attended Iowa State University where she received her Bachelors of Science degree in Agronomy with an emphasis in plant breeding and biotechnology and a minor in Spanish, 2011. While at ISU she was very active in Agronomy Club, Soil Judging Team (placing 1st at a regional contest and 4th at a national contest), Kappa Kappa Psi, Alpha Zeta, and served as a community advisor in the Maple Residence Hall and teaching assistant for Dr. Russell Mullen's class, Principles of Agronomy. She also studied abroad in Panama, Spain, and Costa Rica on Spanish culture and agriculture related trips. In Spain, she interned at the University of Lleida with Dr. Jaume Lloveras and his dry-land agronomy research group, summer 2010. As an undergraduate, Leah worked part-time in a maize breeding laboratory under the direction of Dr. Thomas Lübberstedt for 2 ½ years. She received an undergraduate research fellowship from the Iowa State Agronomy Department. Leah also interned at Pioneer International, Inc. in maize product development in Marion, IA summer, 2009 and interned at Mycogen Seeds (Dow AgroSciences LLC) in maize production in Marshalltown, IA summer, 2008. Leah started a M.S. program in August, 2011 at North Carolina State University in soybean breeding under the guidance of Dr. Thomas Carter, USDA-ARS.

ACKNOWLEDGMENTS

I would like to acknowledge my advisor, Dr. Thomas Carter, for his strong support and dedication as I worked on this thesis. I appreciate so much the discussions and opportunities to learn about plant breeding, agriculture, and research.

I would also like to thank my committee members: Dr. Lilian Miranda, Dr. David Dickey, and Dr. Julian Chaky for all their strong support and guidance.

In addition, I want to recognize the USDA-ARS Soybean and Nitrogen Fixation Unit past and present staff for all the help I received: Dr. Margarita Villagarcia, Cory Callahan, Earl Huie, William Novitzky, Connie Bryant, Dr. Andrew Scaboo, Dr. Joseph Burton, Robert Masson, Dr. Phil Rzewnicki, Hailey Huie, Katherine Novitzky, Jimmy Kanowski, Mackenzie Gibbons, Forrest Howell, Brad Rush, Bobby McMillen, and John Graber. And to the past and present graduate students in soybean breeding at the Soybean and Nitrogen Fixation Unit that supported me through my work on this thesis: Jacob Delheimer, Rochelle Strednak, Amanda Noble, Paul Ruddle, and David Eickholt. Without everyone's help and encouragement this would have never been possible.

Lastly, I would like to thank the USDA-ARS and DuPont Pioneer for their funding and support of this project. Dr. Wendy Srnic for her mentorship through DuPont Pioneer. Dr. Zack Shearin, Michael VanStaalduinen, and Gary Hamm through DuPont Pioneer for their help and collaboration in running a location of my experiment and Dr. Roger Boerma, Dale Wood, and Gina Bishop at the University of Georgia for their help and collaboration running a location of my experiment. Also, Cathy Herring and the station workers at the Central Crop Research Station for their help with my nursery materials in Clayton, NC.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

LITERATURE REVIEW

Need for genetic diversity in the soybean gene pool

Humans depend on crop improvement and additional means to enhance crops, especially with the continued increase in world population. In order to enhance crops, genetic diversity is needed. Soybean [*Glycine max* (L.) Merr.] improvement efforts in North America have produced superior soybean varieties; however, most varieties descend from only a few ancestral lines (Delannay et al., 1983; St Martin, 1982). The breeding procedure in soybean has been one of introduction, hybridization, selection, and hybridization again with parents that are high-yielding and have good agronomic phenotypes to create populations for selection (Li et al., 2008; Luedders, 1977; Paschal and Wilcox, 1975). Intensive selection pressure in soybean populations has ultimately resulted in small yield gains throughout the years and a narrow genetic base between and among superior varieties (Kisha et al., 1997; Manjarrez-Sandoval et al., 1997). The North American soybean gene pool traces back to 50 PIs with about fourteen of those ancestral lines founding over half the North American soybean genetic base (Gizlice et al., 1993). A comparison of the diversity in Asian soybean landraces to North American cultivars which used phenotypic characterization (Cui et al., 2000b; Cui et al., 2001) and coefficient parentage (Cui et al., 2000a) revealed that North American cultivars have a lower level of diversity than Asian landraces. Successive genetic bottlenecks between wild and cultivated soybean and between Asian landraces and North American cultivars also confirmed the reduced genetic diversity through genome sequence analyses (Hyten et al., 2006).

In applied plant breeding, it is commonly thought that the cross of a genetically diverse set of parents rather than the cross of a genetically similar set of parents boosts the level of breeding success (Manjarrez-Sandoval et al., 1997). Breeding in new germplasm from accessions in seed banks such as the USDA soybean collection would broaden the North American soybean genetic base and could provide superior gene combinations (Paschal and Wilcox, 1975). Likewise, infusion of genes from wild soybean genomes could broaden the soybean genetic base.

Use of wild soybean in genetics and breeding

Exploitation of wild relatives and their genetic diversity is essential to guarantee global food security and sustainable crop production (Ratnaparkhe et al., 2011). The use of wild germplasm in breeding programs has been studied in numerous crop species. For example, wild species have been used as a source of resistance to pests in such crops as, tomato (*Solanum lycopersicum* L.), cotton (*Gossypium spp.* L.), and sugarcane (*Saccharum officinarum* L.; Damania, 2008; Harlan, 1976). Yield improvements also have been seen in rice (*Oryza sativa* L.) and wheat (*Triticum aestivum* L.) through wild germplasm utilization. In rice, yield quantitative trait loci (QTL) from the wild rice relative (*Oryza rufipogon* Griff.) were associated with a 17 to 18 percent increase in grain yield (Xiao et al., 1998). Similarly, the increased seed yield in wheat was associated with the genes of the wild wheat relative, *Triticum tauschii* Coss. (MujeebKazi et al., 1996). As a result of the crop improvements made with wild germplasm, crop breeders and researchers have been enticed to further exploit wild genomes in search of valuable traits (Damania, 2008).

In soybean, breeders have exploited two of its wild relatives, *Glycine tomentella* and *Glycine soja* (Siebold and Zucc.) accessions (Kabelka et al., 2004; Singh et al., 1990). Successful introgression of perennial *G. tomentella* (2*n*=78) into *G. max* has been demonstrated. Genetically stable lines were obtained in crosses between *G. tomentella* and *G. max* and these lines showed significantly improved disease resistance to Phytophthora root rot (*Phytophthora sojae*) and sudden death syndrome (*Fusarium virguliforme*) (Ma and Nelson, 2012). Soybean rust (*Phakopsora pachyrhizi*) resistance was also observed in lines with *G. tomentella* in their pedigree (Chang and Hartman, 2010). These examples further exemplify the valuable traits breeders discovered through wild germplasm exploitation.

Even more interesting for soybean breeding, annual *G. soja* is of particular interest for soybean breeding because it is thought to be the wild progenitor of cultivated soybean (Harlan and deWet, 1971). Although *G. max* and *G. soja* are different species by classical taxonomy (Herman, 1962), both species contain *2n*=2x=40 chromosomes and hybridize readily to produce viable fertile offspring (Karasawa, 1936; Singh and Hymowitz, 1988; Weber, 1950). Molecular studies also verify that *G. max* and *G. soja* are genomically similar (Doyle and Beachy, 1985; Kollipara et al., 1995). This similarity and hybridization ability has led soybean breeders to believe that the wild soybean may hold valuable genes for introgression into *G. max*.

The wild soybean includes a range of traits in flower, pubescence, seed and hilum color, disease and insect resistance, and percent protein, oil, and carbohydrates (Carter et al., 2004; Hymowitz and Bernard, 1991; Weber, 1950). *G. soja* may hold a wealth of phenotypic and genetic variability for soybean breeders to exploit; however, it has undesirable traits such

as vining, lodging, seed shattering, and black seed coats (Carpenter and Fehr, 1986). Even with these undesirable characteristics, *G. soja* is of particular interest for breeding because of its rich genetic diversity. Genomic research shows that *G. soja* to be more genetically diverse than *G. max* (Li and Nelson, 2002; Nichols et al., 2007). In a comparison of 192 *G. soja* accessions and 104 *G. max* accessions, eight SSR markers used to associate seed protein content and resistance to Sclerotinia stem rot (*Scleotinia sclerotiorum*) revealed that *G. soja* had a higher number of detectable alleles and a higher level of genetic variability than *G. max*. Furthermore, in this comparison *G. soja*'s average gene diversity was 0.74 and *G. max*'s average gene diversity was 0.64 based on single-locus heterozygosity at the eight SSR marker loci (Jun et al., 2011).

The germplasm collections have been underutilized gene sources for improvement of quantitative traits, such as yield; despite evidence that wild species may have alleles which positively influence agronomic traits (Tanksley and McCouch, 1997). There have been limited soybean breeding efforts to take advantage of the rich reservoir of genetic diversity in the USDA soybean germplasm collection (Singh and Hymowitz, 1999; Carter et al., 2004; and Singh et al., 2007). While the wild soybean germplasm collection has been underutilized, some efforts using *G. soja* are on record (Table 1.1). A high-protein QTL and resistance to soybean cyst nematode was found in *G. soja* (Kabelka et al., 2006; Sebolt et al., 2000). Tolerance to the herbicide, metribuzin [4-Amino-6-(1,1-dimethylethyl)- 3-(methylthio)-1,2,4-triazin-5(4H)-one] was also identified (Kilen and He, 1992). Additionally, wild soybean germplasm has been introduced into a small number of North American breeding programs to develop small seeded specialty cultivars or soyfood

cultivars, such as ‘Canatto’, ‘Nattawa’, and ‘Nattosan’ from Agriculture Canada (Cui et al., 2004; Table 1.1).

Yield improvement using wild soybean

Two yield alleles from *G. soja* accession PI 245331 were found by Li and Pfeiffer (2008). The alleles were identified in a population developed through an advanced backcrossing strategy and were associated with a 6.3% increase in seed yield. One of the yield QTL was found at Satt51 on linkage group A1. The other positive yield allele was identified at Satt529 on linkage group J; this allele was also found to be linked to plant height. Plant height could have confounded the true allele effect as increased seed yield may have resulted from increased plant height (i.e. more pods per plant because of taller plants). Concibido et al. (2003) discovered a *G. soja* allele from PI 407305 associated with a 9.3% increase in seed yield for the recurrent *G. max* parent HS-1 (Jacon Hartz Seed, Stuttgart, Arkansas, USA). The yield QTL from *G. soja* was found on linkage group B2 linked to SSR marker Satt066. Concibido et al. (2003) identified the yield allele in a BC_2 population developed through advanced backcrossing coupled by molecular marker analysis and phenotypic evaluation in the BC_2 population, as described by Tanksley and Nelson (1996). The allele was introgressed into six elite soybean cultivars, some more productive than HS-1. However, little to no effect of the allele was observed in the elite cultivars. A significant ($P \leq 0.05$) positive QTL effect on yield was observed in only two of the six backgrounds. These results suggested that those cultivars have yield alleles similar to the *G. soja* allele. Thus, it would not have been advantageous to introgress the *G. soja* allele into some of these elite germplasm backgrounds. These two studies showed that yield potential exists in *G. soja* and

suggests that other accessions in the wild germplasm collection may hold genes that can positively influence yield. Other studies exist that have attempted to utilize wild soybean, though they had little success. Ertl and Fehr (1985) made *G. max* X *G. soja* crosses followed by five backcrosses to the *G. max* recurrent parent. They reported that even under selection, introgression of *G. soja* into *G. max* was not an effective way to increase yield potential. Carpenter and Fehr (1986) concluded that under no selection, to obtain a high frequency of agronomically acceptable segregates, three backcrosses were needed, which is equivalent to about six percent *G. soja* alleles in derived populations. Chaky et al. (2003) used an advanced backcrossing method to incorporate *G. soja* alleles into a *G. max* background. QTL analyses found a significant ($P \leq 0.05$) seed yield QTL, although the additive effect of the QTL caused a reduction in seed yield.

Heterosis in soybean hybrids

Heterosis interpretations and definitions

When two lines are intercrossed, their progeny hybrids frequently exceed the parental values for several characters. This increased performance in hybrids is called heterosis, otherwise known as hybrid vigor (Fehr, 1993). Heterosis can contribute to characteristics of economic interest, such as drought tolerance, disease resistance, and yield (Duvick, 2001). The term heterosis was first introduced by Shull in 1914 (Shull, 1948) after he conducted maize inbred and crossing experiments. Midparent heterosis (MPH) represents performance of the hybrid that exceeds the average performance of the parents *per se* and is calculated as:

$$\text{Midparent heterosis (\%)} = \frac{(F_1 - \text{MP})}{\text{MP}} \times 100$$

where, F_1 = performance of the hybrid;

MP = average of parents *per se* (parent 1 + parent 2)/2.

High-parent heterosis (HPH) occurs when the hybrid performance exceeds that of the best parent or best performing parent and is calculated as:

$$\text{High-parent heterosis (\%)} = \frac{(F_1 - \text{HP})}{\text{HP}} \times 100$$

where, F_1 = performance of the hybrid;

HP = parent with the highest yield or best parent (Fehr, 1993).

Determining MPH is valuable for understanding theoretical aspects of heterosis and understanding how crossing reverses the effects of inbreeding. In addition, HPH (or better parent heterosis) is predominately more important to demonstrate an economical advantage to growing hybrids crops instead of inbred cultivars (Walsh, 2003).

The genetic theory behind heterosis has been one of debate since the early twentieth century. Heterosis has been interpreted to reflect gene action other than additive (Crow, 2001). What researchers do know is that heterosis is determined by non-mutually exclusive mechanisms, which include dominance, overdominance, and epistasis (Lippman and Zamir, 2007). These three theories are the foundation of most studies that try to explain this biological phenomenon.

The first theory, dominance, assumes that detrimental recessive allele(s) from one parent are masked by dominant allele(s) from the other. Dominance is a non-additive intralocus allelic interaction (Crow, 2000). Falconer and Mackay (1996) explained heterosis in terms of dominance in a quantitative genetics model. In their model, they looked at the

effects of dominance at a single locus. For a single locus with two alleles whose frequencies are p and q in one population and p' and q' in the other population, the difference of gene frequency between the two populations equal to y, so that $y = p - p' = q - q'$. Also, the genotypic values are a, d, $-a$, where a refers to the A_1A_1 genotype, d refers to the A_1A_2 genotype, and $-a$ refers to the A_2A_2 genotype. In this model, it is assumed that the genotypic values are the same in both populations and the epistatic interactions are ignored. The parental means from the two populations, M_{P1} and M_{P2} are:

$$M_{P1} = a(p - q) + 2dpq$$

$$M_{P2} = a(p - y - q - y) + 2d(p - y)(q + y).$$

The midparent value is:

$$MP = \tfrac{1}{2}(M_{P1} + M_{P2}).$$

If the two populations are randomly crossed to produce an F_1, the mean genotypic value of the F_1 is:

$$M_{F1} = a(p^2 - py - q^2 - qy) + d[2pq + y(p - q)].$$

The amount of heterosis in the F_1 is determined by the subtraction of the mean genotypic value of the F_1 and the midparent value:

$$H_{F1} = M_{F1} - MP$$

$$= dy^2.$$

The result of the equation is that heterosis depends on dominance, d. Therefore, loci without dominance (where $d = 0$) do not cause heterosis. When the joint effects of all loci are taken together, and the genotypic values combine additively, heterosis in the F_1 is:

$$H_{F1} = \sum dy^2.$$

Likewise, F_2 heterosis can be calculated. The gene frequency in the F_1 is $(p - ½\, y)$ for one allele and $(q + ½\, y)$ for the other allele. The mean genotypic value of the F_2 is then:

$$M_{F2} = a(p - ½\, y - q - ½\, y) + 2d(p - ½\, y)(q + ½\, y).$$

Thus, the amount of heterosis observed in the F_2 is the difference between the F_2 and the midparent values (Falconer and Mackay, 1996).

$$H_{F2} = M_{F2} - MP$$

$$= ½\, dy^2.$$

From these equations, four conclusions could be drawn:

1.) Heterosis is dependent on directional dominance. For example, if some loci are dominant in a way that they cause a positive yield response, but some loci are dominant in a way that they cause a negative yield response, then the loci effects cancel each other out and a yield response is not observed. Ultimately, heterosis is not observed. Notably, in the absence of heterosis, there is not sufficient grounds to say $d = 0$.

2.) Heterosis is cross specific. Each cross has a different amount of heterosis because the genes by which two lines differ are not the same for all pairs of lines. Different crosses have different values of $\sum dy^2$. Loci that do not display dominant gene action do not cause heterosis, and the amount of heterosis depends on the square of the difference among allele frequencies between the two parental populations. The alleles by which the two lines differ that are crossed to form the F_1 population are not the same for all pairs of lines, and different pairs of lines have different values of $\sum dy^2$, hence, an explanation for specific combining ability.

3.) Heterosis in a cross between two inbred lines is the sum of the dominance deviation *d* at those loci that carry different alleles in the two inbred lines. Highly inbred lines (completely homozygous) differ in gene frequency between them by only 0 or 1, and, if those lines are crossed, heterosis is a function of dominance at the alleles that differ. The greatest level of heterosis occurs when two unrelated inbred lines are crossed because $d \neq 0$ at loci and inbreds carry different alleles, $y = 1$ (Falconer and Mackay, 1996; Lamkey and Edwards, 1999). The importance of the differences in gene frequencies, *y*, between the two entities being crossed is the reason most heterosis is expected in crosses amongst unrelated populations or inbred lines (Walsh, 2003).

4.) Lastly, heterosis in the F_2 is reduced by half when the F_1 is randomly mated, provided epistatic interactions are absent: H_{F1}: dy^2 vs. H_{F2}: ½ dy^2 (Falconer and Mackay, 1996; Lamkey and Edwards, 1999).

The application of the dominance hypothesis, though strong theoretical support for it, reported little success in the early years of agricultural research. For example, early studies in maize used mass selection to try to increase grain yield, but it was ineffective (Hull, 1945; Hull, 1946). These failures led to the emergence of the overdominance hypothesis. However, more recent studies, with improved selection methodologies and better pollen control, show mass selection as an effective way to increase yield in maize and thus, dominance is highly regarded as a main underlying mechanism behind heterosis (Crow, 2000).

The second theory, overdominance, assumes at least at some loci, heterozygotes are superior to either homozygote; and, intralocus allelic interactions at one or more heterozygous genes lead to increased plant vigor. Therefore, hybrids, which are more heterozygous than inbreds, have better performance. Theoretically, in the overdominance model, only a single heterozygous gene is needed to achieve heterosis. Genes showing overdominance are the most noteworthy and sought after from both a fundamental and applied perspective (Crow, 2000). The first example of a single overdominant gene for yield was reported in tomato by Krieger et al. (2010). Krieger et al. found that heterozygosity for tomato loss-of-function alleles of *SINGLE FLOWER TRUSS* increased yield by up to 60 percent. The study suggested that single heterozygous mutations may improve productivity in other agricultural organisms. Ultimately there is little evidence for overdominance at the population level, but as observed in elite crosses, hybrids may get an extra boost from some overdominant loci (Crow, 2000).

The third theory, epistasis, is defined as a non-additive interlocus allelic interaction. Epistasis assumes that alleles form unique epistatic combinations of additive and dominant alleles that can mask the effect of another allele, which would allow for the observance of heterotic increases in yield (Crow, 2000). It has been proposed that alleles in hybrids are involved in novel allelic interactions which modify the levels of gene expression and that the variation for the presence or absence of genes contributes to the observed heterotic phenotypes (Springer and Stupar, 2007). Presence and absence variation has been studied among maize inbred lines. The studies showed that some genes were present in some maize inbred lines while other inbred lines were absent of those particular genes. The evidence was

strongly supported by observations of crosses made between inbred maize lines in different heterotic groups. Ultimately when two inbred lines from different heterotic groups were crossed, all the genes present together caused epistatic interactions, which contributed to heterosis in maize (Swanson-Wagner et al., 2006). Additional evidence for this view has been seen in polyploids. Progressively more heterosis occurred in polyploids as the diversity of the component genomes increased (Swanson-Wagner et al., 2006). Much of the quantitative genetic theory is based on single locus theory, which assumes the absence of epistasis. From the non-epistatic standpoint, additive and dominance effects account for the performance of hybrids (Falconer and Mackay, 1996). However, it is widely observed that hybrids get an extra boost from some underlying genetic interactions, which is likely due to epistasis (Crow, 2000).

In addition to dominance, overdominance, and epistasis, physiological and biochemical processes could explain heterosis. One biochemical explanation proposed that the heterozygosity of a hybrid relaxed the strict regulatory factors that control plant growth mechanisms. The relaxed state allowed a larger amount of growth in the plant (Milborrow, 1998). Researchers also hypothesized that differential gene expression in inbreds and hybrids were responsible for heterosis. For example, a hybrid accumulated levels of transcript equal to the midparent (additivity), the high or low-parent (high or low-parent dominance), above the high-parent (overdominance), or below the low-parent (underdominance; Swanson-Wagner et al., 2006). Another hypothesis suggested the difference in the performance of hybrids versus inbreds was caused by general multigenic heterosis. Hybrids grew more efficiently than inbreds because they metabolized fewer

specific unstable proteins. Smaller levels of protein metabolism in hybrids were proposed to result from more allelic choices in the heterozygous state. Protein metabolism, an energy demanding process, caused less energy to be available for inbreds, which resulted in less cell growth and growth of a plant. Higher levels of unstable proteins in inbreds also reduced cell cycle progression, which limited growth (Goff, 2011).

Some hybrid combinations bring together factors that cause a stronger restrictive influence (subtractive heterosis) when alleles are present together than when they are separate. This phenomenon is suggested to be missed in most studies due to the emphasis placed on the positive role of heterosis. However, if a negative influence on heterosis did exist in a study, it probably was discounted and considered to be caused by a disease or environmental conditions (Crow, 1998).

Heterosis found previously in *G. max* x *G. max* crosses

Heterosis in soybean has been rarely studied because it is generally thought that little heterosis exists in hybrid soybeans. In addition, it is very difficult and very time-consuming to obtain large quantities of hybrid seed for standard yield plot testing. Many early F_1 heterosis tests produced too little F_1 seed to plant large enough plots for standard testing. Studies of single F_1 plants showed F_1 heterosis ranging from +17 to +25% average MPH and +4 to +13% average HPH (Table 1.2). Reliable yield estimates are difficult to measure from single plant hills or un-bordered single rows. Therefore it is difficult to obtain accurate heterosis estimates and make inferences to commercial production.

The hybrid soybean literature documented evidence of heterosis with the use of standard yield plot techniques in the F_1 and F_2 generation (Burton and Brownie, 2006;

Palmer et al., 2001). Across several F_1 heterosis studies that used standard yield plots, the reported heterosis ranged from -48 to +28% average MPH. The average HPH ranged from -52 to +20% (Table 1.2). A summary of 14 heterosis reports from 1930 to 2001, which encompassed 456 different crosses complied by Palmer et al. 2001, showed the average F_1 and F_2 MPH ranged from +14 to +46% and HPH ranged from +4 to +34%. A recent study by Perez et al. (2009) used standard plot technique and observed positive and negative heterosis. Perez et al. (2009) employed male-sterile female-fertile lines with *ms3* and *ms9* mutations and insect vectors to facilitate the transfer of pollen for hybrid seed production. In 2007, MPH ranged from -29% to +32% and HPH ranged from -32% to +1% from eleven F_1 crosses. In the next year, 2008, the same F_1's were used and MPH ranged from -53% to -21% and HPH from -66% to -35% (Table 1.2). A lack of consistency across years due to a strong genotype by year effect made these results difficult to interpret. Significant ($P \leq 0.05$) positive MPH and HPH for plant height, lodging, seed size, and seed protein existed in several crosses in some years (Perez et al. 2009). Many of these traits are associated with vegetative growth. It is likely that vegetative heterosis may increase early lodging and pod abortion which can reduce seed yield (Lewers et al., 1998a).

In soybean, it is suggested that the possible genetic basis for heterosis could be due to a few effects: complementary gene interaction, a greater number of dominant alleles in the hybrid than in the parent, and inheritance of linked dominant alleles in linkage blocks. Linkage blocks complement each other because they mask less favorable alleles or because there is interaction of duplicate loci. Furthermore, multiple dosage-dependent regulatory loci and/or overdominance have been suggested as a base for heterosis (Burton and Brownie,

2006). Dominance would not be expected in an autogamous species like soybean. Soybeans outcross at a very low frequency, they take on a small genetic load, and are highly inbred (Burton, 2011). Nonetheless, studies show statistically significant heterosis, which is clear evidence that dominance is important in some breeding populations (Brim and Cockerham, 1961; Burton and Brownie, 2006).

Correlation between heterosis and genetic relationship and heterosis prediction

Cultivated crop plants have provided the genetic tools to study heterosis because parental inbreds have been artificially selected for maximum combining ability. The creation of structured genetic populations enables more accurate quantitative phenotyping than in natural populations (Lamkey and Edwards, 1999). As plant breeders continue to improve inbreds and to define the most superior parental combinations, many unanswered questions still exist about the genetic basis of heterosis. The main questions revolve around what genes are involved. Heterosis is usually associated with the concept of genetic divergence for the maximum expression of traits. Genetic diversity has been suggested as the key to obtain hybrid vigor and that the greater the diversity of the parental lines, the higher yielding the progeny (Moll et al., 1965). The assumption that heterosis increases with the diversity of the uniting gametes goes back to Shull (1908) and East (1908). It was assumed that there was a physiological stimulus to development which increased with more genetic dissimilarity and could be observed in wide crosses. Though an increase in the degree of genetic differentiation between the two populations is important to capture heterosis, the differences between isolated populations may be too great and may cause interspecific sterility and a decrease in heterotic expression (Walsh, 2003). Based on the assumption that greater

heterosis is indicative of a greater degree of genetic differentiation between parents, the performance of F_1 or F_2 hybrids could be used to predict hybrid or inbred line performance (Weiss et al., 1947). The attempts to relate yield heterosis to parental genetic distance or co-ancestry has produced mixed results. In maize (*Zea mays* L.) and sunflower (*Helianthus annuus* L.) there was evidence that hybrid performance can be correlated to genetic distance (Lee et al. 1989; Smith et al., 1990; Lanza et al., 1997; Cheres et al., 2000). However, there was other evidence in maize that showed genetic distance poorly correlated to heterosis when simple sequence repeat (SSR) markers and amplified fragment length polymorphisms (AFLP) were used to determine genetic distance (Guimaraes et al., 2007). In cotton (*Gossypium hirsutum* L.), genotypes with the maximum range of genetic distance were better predictors of hybrid performance for some traits (Gutiérrez et al., 2002). In a study of heterosis in dry edible bean (*Phaseolus vulgaris* L.) and faba bean (*Vicia faba* L.), yield was highly correlated to parental genetic distance for dry bean, but little association was found in faba bean (Ghaderi et al., 1984). In soybean, little relationship between seed yield heterosis in the F_1 and genetic distance was found (Cerna et al., 1997; Paschal and Wilcox, 1975). In addition, genetic distance was found to be a poor predictor of overall F_2 performance (Gizlice et al., 1993; Manjarrez-Sandoval et al., 1997; Nelson and Bernard, 1984). These reported studies used only a few markers to determine genetic distance of parental lines. With the advanced technology today, the use of a large number of single-nucleotide polymorphisms (SNP) to determine genetic distance might change the picture of the correlation between hybrid performance and parental genetic distance. In most of the reported heterosis studies, cultivated soybean has been the only parental stock in the development of hybrids. The few

studies that look at *G. max* x *G. soja* crosses have not used standard yield testing plot techniques and have not measured heterosis *per se* (William, 1948; Weber, 1950). While many works suggest genetic dissimilarity does not have any correlation to hybrid vigor, association of parental divergence and progeny performance in crosses with wild soybean in the parental pedigree has not been documented. Since *G. soja* is more genetically dissimilar to *G. max* (Jun et al., 2011; Nichols et al., 2007), hybrids of breeding lines with some *G. soja* alleles in their background may have the potential for significant heterosis that can be correlated to and predicted by genetic distance, co-ancestry, or percent *G. soja* alleles.

F_2 heterosis as an inbred line predictor

Heterosis has been observed in soybean, indicating that it may be worthwhile to search for favorable heterozygous gene combinations (Brim, 1973). For most traits, the majority of genotypic variation is controlled by additive genetic variance. Non-additive genetic effects, such as dominance and epistatic variation, also control traits, but it is generally smaller than additive variation. In pure-line cultivar development, dominance and additive X additive epistasis can be fixed with inbreeding. Single locus interactions – overdominance, and additive X dominance and dominance X dominance types of epistasis are not fixable in homozygotes (Burton, 2012). Therefore, if positive heterosis is a result of beneficial dominant loci, then superior gene combinations can also be fixed. However, heterosis may not adequately reflect non-additive gene action for a self-fertilizing species like soybean, since dominance and epistatic deviations may be both plus and minus and will tend to cancel out when averaged over loci (Brim, 1973).

Typically not enough F_1 soybean seeds are produced for standard yield trials. Single hill plot tests and single row tests of F_1 seed often produce inflated F_1 yield values, which are unreliable yield estimates to a breeder. Alternatively, if the hybrids are allowed to self-pollinate one generation, sufficient amounts of F_2 seed can be produced and bulked together for standard yield trials (Burton and Brownie, 2006). Although not all alleles are in a heterozygous state in the F_2 generation, at least half of the genes are (Fehr, 1993). Heterosis in the F_2 can be explored as there is much evidence of heterosis in the F_2 generation (Table 1.2). For a plant breeder, the evidence of F_2 heterosis could be predictive of good parental combinations (Burton and Brownie, 2006). It would also be preliminary evidence that superior gene combinations are possible and new/unique advantageous yield alleles may exist in the breeding material (Scaboo and Carter, 2012). If the F_2 bulk populations are predictive of inbred line performance, intercross selection with yield tests of F_2 bulks would be advantageous (Friedrichs, 2009). The measured F_2 heterosis could be used as a prioritization tool for PIs when incorporating new germplasm into a breeding program. The lines that have the best heterosis should be the lines that receive the most focus to incorporate potentially new yield alleles into a breeding program (Scaboo and Carter 2012). When additive X additive effects are important, early generation testing may not be an appropriate selection approach since an opportunity must be provided for unique gene combinations to come together. Progress from selection will be slower when additive variance predominates for traits of interest. However with the use of heterosis and with the assumption that it depends on dominance, soybean breeders could advance and enhance their selections with heterosis knowledge (Brim, 1973; Burton, 2011).

Interpretation of F_2 heterosis in *G. max* x *G. soja* crosses

The wild progenitor of cultivated soybean, *G. soja*, has many undesirable characteristics in comparison to modern adapted cultivars (Carpenter and Fehr, 1986). The few studies that reported on the performance of hybrids made between direct crosses of adapted and wild soybean, revealed that the undesirable wild soybean traits were still prominent in the F_1, F_2, and BC_1 generations. The progenies in these generations had poor appearance and performance, were inadaptable, and the erect habit of the domesticate parent was not recovered (Williams, 1948; Weber, 1950; Delheimer, 2012). Direct crosses of adapted soybean to wild soybean to measure F_1 or F_2 heterosis for inbred line prediction would be unfavorable. Weber (1950) reported seed size data on F_1 and F_2 single plants from direct *G. max* x *G. soja* crosses and the calculated heterosis values were -35% MPH in the F_1 and -45% MPH in the F_2 in 1940, with similar results obtained in 1943. Resulting F_1 or F_2 heterosis in this type of cross were highly influenced by the poor undesirable traits of the wild soybean. If the wild progenitor had beneficial genes, such as yield genes, the unadaptedness of the germplasm would mask the effect of beneficial alleles. The decrease in the population mean is proportional to the overall unadaptedness of the germplasm containing the superior alleles (Murphy, 2011). Through intense selection, however, adapted inbred lines with a moderate percent of wild soybean alleles can be derived from a *G. max* x *G. soja* cross (Delheimer, 2012). Adapted lines can be selected for desirable characteristics and also high yield, while maintaining wild soybean alleles in the genetic background. Heterosis measured in F_2 hybrids of crosses made between *G. max* and the *G.*

max x *G. soja* derived inbred lines may be useful to understand the utility of the *G. soja* alleles.

By definition, quantitative traits such as yield are controlled by many loci; so the genetic effects of interactions between alleles at these different loci and how they impact genotypic values and quantitative variation should be considered (Brim, 1973). In terms of dominance contributing to yield heterosis, the alleles in a positive partial dominance model show that the F_1 value will fall between the MP and the HP (Figure 1.1). With positive complete dominance the F_1 will equal that of the HP (Figure 1.2). In the absence of dominance, the hybrid mean is equivalent to the MP genotypic value (Falconer and Mackay, 1996; Figure 1.3). Weber's (1950) data revealed a positive skewness in the distribution of the F_2 seed size in comparison to the F_1 mean . Weber suggested that this skewness could be interpreted as partial dominance of genetic factors in determining small seed size.

REFERENCES

Brim, C.A. 1973. Quantitative genetics and breeding. p. 155-183. In B.E. Caldwell (ed.) Soybeans: Improvement, production, and uses. Agronomy 16 ed. American Society of Agronomy, Inc., Madison, Wisconsin, USA.

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

Burton, J.W. 2012. Quantitative genetics in soybean: Is dominance important? North Carolina State University, Raleigh, NC. Unpublished Manuscript.

Burton, J.W. 2011. "Heterosis in soybean: Can it be useful in breeding?". Crop Science 745 Quantitative Genetic in Plant Breeding Course. North Carolina State University. February 10. http://soybase.org/meeting_presentations/soybean_breeders_ workshop/SBW_2011/ Burton.pdf.

Burton, J.W., and C. Brownie. 2006. Heterosis and inbreeding depression in two soybean single crosses. Crop Sci. 46:2643-2648.

Carpenter, J.A., and W.R. Fehr. 1986. Genetic-variability for desirable agronomic traits in populations containing *Glycine soja* germplasm. Crop Sci. 26:681-686.

Carter, T.E., R.L. Nelson, C.H. Sneller and Z.L. Cui. 2004. Genetic diversity in soybean. p. 303–416. In H.R. Boerma, and J.E. Specht (eds.) Soybeans: Improvement, production, and uses. Agronomy Monographs 3rd ed. No. 16 ed. ASA-CSSA-SSSA, Madison, WI, USA.

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

Chaky, J.M. 2003. Advanced backcross QTL analysis in a mating between *Glycine max* and *Glycine soja*. M.S. thesis. University of Nebraska, Lincoln.

Chang, S., and G. Hartman. 2010. Characterizing soybean rust resistance in *Glycine tomentella* and effect of temperature and a fungicide on *Phakopsora pachyrhizi* uredinia production of urediniospores. M.S. thesis. University of Illinois at Urbana-Champaign.

Cheres, M.T., J.F. Miller, J.R. Crane, and S.J. Knapp. 2000. Genetic distance as a predictor of heterosis and hybrid performance within and between heterotic groups in sunflower. Theor. Appl. Genet. 100:889-894.

Concibido, V.C., B. La Vallee, P. Mclaird, N. Pineda, J. Meyer, L. Hummel, J. Yang, K. Wu, and X. Delannay. 2003. Introgression of a quantitative trait locus for yield from *Glycine soja* into commercial soybean cultivars. Theor. Appl. Genet. 106:575-582.

Crow, J.F. 2000. The rise and fall of overdominance. p. 225-257. In Plant breeding reviews.

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

Cui, Z., A.T. James, S. Miyazaki, R.F. Wilson, and T.E. Carter Jr. 2004. Breeding specialty soybeans for traditional and new soyfoods. In Soybeans as Functional Foods and Ingredients, AOCS Press, Champaign, IL264-322.

Cui, Z.L., T.E. Carter, and J.W. Burton. 2000a. Genetic diversity patterns in Chinese soybean cultivars based on coefficient of parentage. Crop Sci. 40:1780-1793.

Cui, Z.L., T.E. Carter, and J.W. Burton. 2000b. Genetic base of 651 Chinese soybean cultivars released during 1923 to 1995. Crop Sci. 40:1470-1481.

Cui, Z.L., T.E. Carter, J.W. Burton, and R. Wells. 2001. Phenotypic diversity of modern Chinese and North American soybean cultivars. Crop Sci. 41:1954-1967.

Damania, A.B. 2008. History, achievements, and current status of genetic resources conservation. Agron. J. 100:9-21.

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

Doyle, J.J., and R.N. Beachy. 1985. Ribosomal gene variation in soybean (*Glycine*) and its relatives. Theor. Appl. Genet. 70:369-376.

Duvick, D.N. 2001. Biotechnology in the 1930s: The development of hybrid maize. Nature Reviews Genetics 2.

East, E.M. 1908. Hybridization methods in corn breeding. Rep. Am. Breed. Assoc. 6:63-72.

Ertl, D.S., and W.R. Fehr. 1985. Agronomic performance of soybean genotypes from *Glycine max* X *Glycine soja* crosses. Crop Sci. 25:589-592.

Falconer, D.S., and T.F.C. Mackay. 1996. Introduction to quantitative genetics. Longman, Harlow, England.

Fehr, W.R. 1993. Principles of cultivar development. Theory and technique. Macmillan Publishing Company, New York. 519 p.

Friedrichs, M. 2009. Heterosis and inbreeding depression in *Glycine max*, PhD diss., North Carolina State University, Raleigh, NC.

Ghaderi, A., M.W. Adams, and A.M. Nassib. 1984. Relationship between genetic-distance and heterosis for yield and morphological traits in dry edible bean and faba bean. Crop Sci. 24:37-42.

Gizlice, Z., T.E. Carter, and J.W. Burton. 1993. Genetic diversity in North American soybean. 2. Prediction of heterosis in F_2 populations of southern founding stock using genetic similarity measures. Crop Sci. 33:620-626.

Goff, S.A. 2011. A unifying theory for general multigenic heterosis: Energy efficiency, protein metabolism, and implications for molecular breeding. New Phytol. 189:923-937.

Guimaraes, P.d.S., Guidetti Zagatto Paterniani,Maria Elisa Ayres, R.R. Lueders, A.P. de Souza, P.R. Laborda, and K.M. Oliveira. 2007. Correlation between the heterosis of maize hybrids and genetic divergence among lines. Pesquisa Agropecuaria Brasileira 42:811-816.

Gutierrez, O.A., S. Basu, S. Saha, J.N. Jenkins, D.B. Shoemaker, C.L. Cheatham, and J.C. McCarty. 2002. Genetic distance among selected cotton genotypes and its relationship with F_2 performance. Crop Sci. 42:1841-1847.

Harlan, J.R., and de Wet, J. M. J. 1971. Toward a rational classification of cultivated plants. Taxon 20:509-519.

Harlan, J.R. 1976. Genetic resources in wild relatives of crops. Crop Sci. 16:329-333.

Herman, F.J. 1962. A revision of genus *Glycine* and its immediate allies. USDA Tech Bull1-79.

Hillsman, K.J., and H.W. Carter. 1981. Performance of F_1 hybrid soybeans in replicated row trials. Agronomy Abstracts. American Society of Agronomy, Madison, WI, p. 63.

Hull, F.H. 1946. Overdominance and corn breeding where hybrid seed is not feasible. Journal of the American Society of Agronomy 38(12).

Hull, F.H. 1945. Recurrent selection for specific combining ability in corn. Journal of the American Society of Agronomy 37(2).

Hymowitz, T., and R.L. Bernard. 1991. Origin of the soybean and germplasm introduction and development in North America. Shands, H.L.and L.E.Wiesner (Ed.).CSSA Special Publication, no.17.use of Plant Introductions in Cultivar Development, Part 1; Symposium, Las Vegas, NV, USA, October 19, 1989147-164.

Hyten, D.L., Q. Song, Y. Zhu, I. Choi, R.L. Nelson, J.M. Costa, J.E. Specht, R.C. Shoemaker, and P.B. Cregan. 2006. Impacts of genetic bottlenecks on soybean genome diversity. Proc. Natl. Acad. Sci. U. S. A. 103:16666-16671.

Johnson, S.L., W.R. Fehr, G.A., Welke, and S.R., Cianzio. 2001. Genetic Variability for Seed Size of Two- and Three-Parent Soybean Populations. Crop Sci. 41: 1029–1033.

Jun, T., K. Van, M.Y. Kim, M. Kwak, and S. Lee. 2011. Uncovering signatures of selection in the soybean genome using SSR diversity near QTLs of agronomic importance. Genes Genom. 33:391-397.

Kabelka, E.A., S.R. Carlson, and B.W. Diers. 2006. *Glycine soja* PI 468916 SCN resistance loci's associated effects on soybean seed yield and other agronomic traits. Crop Sci. 46:622-629.

Kabelka, E.A., B.W. Diers, W.R. Fehr, A.R. LeRoy, I.C. Baianu, T. You, D.J. Neece, and R.L. Nelson. 2004. Putative alleles for increased yield from soybean plant introductions. Crop Sci. 44:784-791.

Karasawa, K. 1936. Crossing experiments with *G. max* and *G. ussuriensis*. Japanese Journal of Botany 8:113-118.

Kaw, R.N., and P.M. Menon. 1979. Heterosis in a 10-parent diallel cross in soybean. Indian J. Agric. Sci. 49:322-324.

Kilen, T.C., and G.H. He. 1992. Identification and inheritance of metribuzin tolerance in wild soybean. Crop Sci. 32:684-685.

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

Kollipara, K.P., R.J. Singh, and T. Hymowitz. 1995. Genomic relationships in the genus *Glycine* (Fabaceae, Phaseoleae) - use of a monoclonal-antibody to the soybean bowman-birk inhibitor as a genome marker. Am. J. Bot. 82:1104-1111.

Krieger, U., Z.B. Lippman, and D. Zamir. 2010. The flowering gene *SINGLE FLOWER TRUSS* drives heterosis for yield in tomato. Nat. Genet. 42.

Lamkey, K.R., and J.W. Edwards. 1999. The quantitative genetics of heterosis. In The quantitative genetics of heterosis. J.G. Coors and S. Pandey (ed.) proceedings of the international symposium on the genetics and exploitation of heterosis in crops, CIMMYT, Mexico City, Mexico, 17-22 Aug. 1997. ASA, CSSA, and SSSA, Madison, WI. 1999.

Lanza, L., C. de Souza Jr., L. Ottoboni, M. Vieira, and A.P. de Souza. 1997. Genetic distance of inbred lines and prediction of maize single-cross performance using RAPD markers. Theoretical and Applied Genetics 94:1023-1030.

Lee, M., E.B. Godshalk, K.R. Lamkey, and W.W. Woodman. 1989. Association of restriction fragment length polymorphisms among maize inbreds with agronomic performance of their crosses. Crop Sci. 29:1067-1071.

Leffel, R.C., and M.G. Weiss. 1958. Analysis of diallel crosses among ten varieties of soybeans. Agron. J. 50:528-534.

LeRoy, A., W. Fehr, and S. Cianzio. 1991. Introgression of genes for small seed size from *Glycine soja* into *G. max*. Crop Sci. 31:693-697.

Lewers, K.S., S.K. St Martin, B.R. Hedges, and R.G. Palmer. 1998a. Effects of the *Dt2* and *S* alleles on agronomic traits of F_1 hybrid soybean. Crop Sci. 38.

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

Li, D., T.W. Pfeiffer, and P.L. Cornelius. 2008. Soybean QTL for yield and yield components associated with *Glycine soja* alleles. Crop Sci. 48:571-581.

Li, Z.L., and R.L. Nelson. 2002. RAPD marker diversity among cultivated and wild soybean accessions from four Chinese provinces. Crop Sci. 42:1737-1744.

Lippman, Z.B., and D. Zamir. 2007. Heterosis: Revisiting the magic. Trends in Genetics 23.

Loiselle, F., H.D. Voldeng, P. Turcotte, and C.A. Stpierre. 1990. Analysis of agronomic characters for an 11-parent diallel of early-maturing soybean genotypes in eastern Canada. Canadian Journal of Plant Science 70:107-115.

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

Ma, J., and R. Nelson. 2012. Evaluation of fertile lines derived from the hybridization of *Glycine max* and *G. tomentella* (2*n*=78) is a wild, perennial relative. M.S. thesis, University of Illinois at Urbana-Champaign.

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

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

Moll, R.H., Lonnquis.Jh, J.V. Fortuno, and E.C. Johnson. 1965. Relationship of heterosis and genetic divergence in maize. Genetics 52:139-&.

MujeebKazi, A., V. Rosas, and S. Roldan. 1996. Conservation of the genetic variation of *triticum tauschii* (coss) schmalh (*aegilops squarrosa auct* non L) in synthetic hexaploid wheats (*T-turgidum* L s lat x *T-tauschii*; 2n=6x=42, AABBDD) and its potential utilization for wheat improvement. Genet. Resour. Crop Evol. 43:129-134.

Murphy, P. 2011. Plant Breeding Methods. Dept. of Crop Science. North Carolina State University. Course notes CS, HS 541.

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

Nichols, D.M., K.D. Glover, S. Carlson, J. Specht, and B. Diers. 2006. Fine mapping of a seed protein QTL on soybean linkage group I and its correlated effects on agronomic traits. Crop Sci. 46:834-839.

Nichols, D.M., W. Lianzheng, Y. Pei, K.D. Glover, and B.W. Diers. 2007. Variability among Chinese *Glycine soja* and Chinese and North American soybean genotypes. Crop Sci. 47:1289-1298.

Ortiz-Perez, E., S.R. Cianzio, H. Wiley, H.T. Horner, W.H. Davis, and R.G. Palmer. 2007. Insect-mediated cross-pollination in soybean [*Glycine max* (L.) Merrill] I. agronomic performance. Field Crops Res. 101:259-268.

Palmer, R.G., J. Gai, H. Sun and J.W. Burton. 2001. Production and evaluation of hybrid soybean. In Plant Breed. Rev. (ed.). p. 263-307.

Paschal, E.H., and J.R. Wilcox. 1975. Heterosis and combining ability in exotic soybean germplasm. Crop Sci. 15:344-349.

Perez, P.T., S.R. Cianzio, and R.G. Palmer. 2009. Evaluation of soybean [*Glycine max* (L.) Merr.] F_1 hybrids. Journal of Crop Improvement 23:1-18.

Ratnaparkhe, M.B., R.J. Singh and J.J. Doyle. 2011. Glycine. p. 83-116. In C. Kole (ed.) Wild crop relative: Genome and breeding resources: Legume crops and forages.

Scaboo, A.M., and T.E.Carter. 2012. Developing and distributing exotic breeding lines by estimating hybrid vigor in the F_2 generation. Abstract In Southern Soybean Breeders' Tour, Delta Research Center, Portageville, MO.

Sebolt, A.M., R.C. Shoemaker, and B.W. Diers. 2000. Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. Crop Sci. 40:1438-1444.

Shull, G.H. 1948. What is heterosis. Genetics 33:439-446.

Shull, G.H. 1908. The composition of a field of maize. Rep. Am. Breeders' Assoc. 4:296-301.

Singh, R.J., and T. Hymowitz. 1999. Soybean genetic resources and crop improvement. Genome 42:605-616.

Singh, R.J., and T. Hymowitz. 1988. The genomic relationship between *Glycine max* (L) Merr and *Glycine soja* Sieb. and Zucc. as revealed by pachytene chromosome analysis. Theor. Appl. Genet. 76:705-711.

Singh, R.J., G.H. Chung, and R.L. Nelson. 2007. Landmark research in legumes. Genome 50:525-537.

Singh, R.J., K.P. Kollipara, and T. Hymowitz. 1990. Backcross-derived progeny from soybean and *Glycine-tomentella* hayata intersubgeneric hybrids. Crop Sci. 30:871-874.

Smith, O.S., J.S.C. Smith, S.L. Bowen, R.A. Tenborg, and S.J. Wall. 1990. Similarities among a group of elite maize inbreds as measured by pedigree, F_1 grain-yield, grain-yield, heterosis, and RFLPs. Theor. Appl. Genet. 80:833-840.

Springer, N.M., and R.M. Stupar. 2007. Allelic variation and heterosis in maize: How do two halves make more than a whole? Genome Res. 17:264-275.

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

Swanson-Wagner, R.A., Y. Jia, R. DeCook, L.A. Borsuk, D. Nettleton, and P.S. Schnable. 2006. All possible modes of gene action are observed in a global comparison of gene expression in a maize F_1 hybrid and its inbred parents. Proc. Natl. Acad. Sci. U. S. A. 103.

Tanksley, S.D., and S.R. McCouch. 1997. Seed banks and molecular maps: Unlocking genetic potential from the wild. Science 277:1063-1066.

Tanksley, S.D., and J.C. Nelson. 1996. Advanced backcross QTL analysis: A method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor. Appl. Genet. 92:191-203.

Veatch, C. 1930. Vigor in soybeans as affected by hybridity. J. Am. Soc. Agron. 22:289-310.

Walsh, B. 2003. Lecture 6: Inbreeding and heterosis. Southern Institute in Statistical Genetics, Melbourne.

Weber, C.R. 1950. Inheritance and interrelation of some agronomic and chemical characters in an interspecific cross in soybeans, *Glycine max* x *G. ussuriensis*. Iowa Agric Expt Sta Res Bull 374:765-816.

Weber, C.R., L.T. Empig, and J.C. Thorne. 1970. Heterotic performance and combining ability of 2-way F_1 soybean hybrids. Crop Sci. 10:159-&.

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

Williams, L. 1948. Inheritance In A Species Cross In The Soybean. P. 131-132. Abstract In 1948. Genetics, Vol 33.

Xiao, J.H., J.M. Li, S. Grandillo, S.N. Ahn, L.P. Yuan, S.D. Tanksley, and S.R. McCouch. 1998. Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. Genetics 150:899-909.

TABLES

Table 1.1 *G. soja* plant introductions (PI) with traits found and the percent *G. soja* alleles or number of backcrosses in the derived lines studied.

Plant Introduction	Trait	Percent *G. Soja* Alleles or Number of Backcrosses	References
PI 468916	High-protein QTL and high protein lines	BC_3F_4	Sebolt et al. (2000), Nichols et al. (2006)
PI 163453	Tolerance to herbicide, metribuzin [4-Amino-6-(1,1-dimethylethyl)- 3-(methylthio)-1,2,4-triazin-5(4H)-one]	F_1	Kilen and He (1992)
PI 245331	Tolerance to herbicide, metribuzin [4-Amino-6-(1,1-dimethylethyl)- 3-(methylthio)-1,2,4-triazin-5(4H)-one]	F_1	Kilen and He (1992)
PI 468916	Resistance to soybean cyst nematode, 2 QTL	BC_4F_3	Kabelka et al. (2006)
PI 101404B	Small seed size trait	12.5 - 25%	Cui et al. (2004)
PI 440913B	Small seed size trait	BC_3F_2	LeRoy et al. (1991)
PI 81762	Small seed size trait	BC_3F_2	LeRoy et al. (1991), Cui et al. (2004)
PI 424004 A	Small seed size trait	BC_3F_2	LeRoy et al. (1991)
PI 135624	Small seed size trait	.	Johnson et al. (2001), Cui et al. (2004)
PI 245331	Two yield QTL	BC_2F_4	Li and Pfeiffer (2008)
PI 407305	One yield QTL	BC_2	Concibido et al. (2003)

Table 1.2 Percent midparent heterosis (MPH) and high-parent heterosis (HPH) in the F_1 and F_2 soybean populations observed in studies conducted with non-standard and standard yield plots.

		Heterosis for Yield						Trial				
F_1	F_2	Non-Standard Yield Plots[#]	Standard Yield Plots[§]	Range of Average MPH%	Range of Average HPH%	Average MPH%	Average HPH%	Number of Crosses	No. Yrs	No. Locs	No. Reps	Reference
X		X		---	---	+36	+20	16	1	1	1	Veatch (1930)
X		X		---	---	+21	+15	17	1	1	1	Weiss et al. (1947)
X		X		---	---	+14	---	45	1	1	2	Leffel and Weiss (1958)
X		X		---	-39 to +90	+25	+13	85	1	1	1	Weber et al. (1970)
X		X		---	---	---	+8	30	2	1	5	Paschal and Wilcox (1975)
X		X		-26 to +104	-33 to +57	+17	+4	45	1	1	3	Kaw and Menon (1979)
X			X¶	---	---	+28	+20	2	2	2	2	Brim and Cockerham (1961)
X			X	---	---	+13	+6	8	2	1	3	Hillsman and Carter (1981)
X			X	-15 to +16	---	+8	+3	27	2	2	3	Nelson and Bernard (1984)
X			X	NR† to +16	---	+5	---	36	2	3	3	Lewers et al. (1998b)
X			X	---	---	+13	+11	2	2	3	3	Burton and Brownie (2006)
X			X	-34 to +15	-41 to +11	-1	-11	9	1	3	5	Ortiz-Perez et al. (2007)
X			X	- 29 to +32	-23 to +1	-6	-22	12	1	3	2	Perez et al. (2009) [2005 trial]
X			X	-53 to -21	-66 to -35	-48	-52	12	1	2	2	Perez et al. (2009) [2006 trial]
	X		X	---	---	+11	---	55	1	3	3	Loiselle et al. (1990)
	X		X	0 to +15	-3 to +11	+7	+3	24	1	2	8	Manjarrez-Sandoval et al. (1997)
	X		X	---	---	+9	+7	2	2	3	3	Burton and Brownie (2006)
	X		X	-1 to +19	---	+9	---	10	2	2	3	Gizlice et al. (1993)

[#]Hill plots, spaced plants, or 1 row un-bordered plots.

[§]Standard 3 or 4 row yield plots.

†NR = not reported because non-significant

¶2 row plot

FIGURES

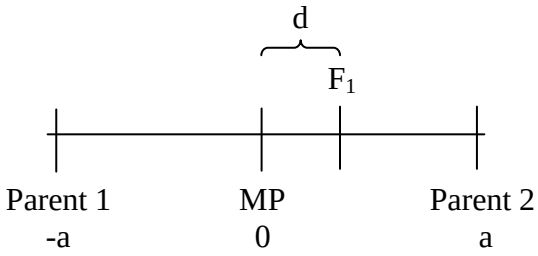


Figure 1.1 Expected F_1 genotypic value in the presence of partial dominance.

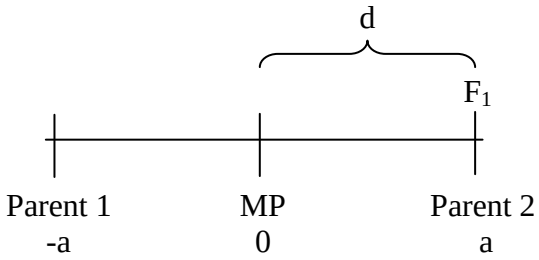


Figure 1.2 Expected F_1 genotypic value in the presence of complete dominance.

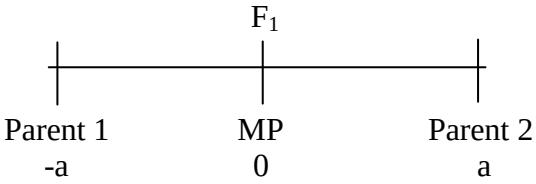


Figure 1.3 Expected F_1 genotypic value in the absence of dominance, epitasis, and overdominance.

CHAPTER TWO

HETEROSIS AS EVIDENCE OF YIELD ALLELES FROM WILD SOYBEAN

INTRODUCTION

The USDA soybean germplasm collection has been vastly underutilized in the improvement of soybean (*Glycine max* L. Merr.) yield. More than 80% of the gene pool in the United States originates from less than a dozen soybean introductions, implying that the genetic base is likely too narrow for sustained rapid yield advances in commercial soybean breeding (Delannay et al., 1983; Gizlice et al., 1993). To guarantee future global food security and sustainable crop production, there is a strong need to broaden the genetic base of soybean cultivars. The wild relatives of soybean are a potential reservoir of diversity for this purpose. At present, the perennial *Glycine tomentella* (Hayata) and the annual *Glycine soja* (Sieb. and Zucc.) have been hybridized successfully with the domesticated soybean to produce breeding lines suitable for yield testing (Ma and Nelson, 2012; Kabelka et al., 2004; Singh et al., 1990). Of the two, the wild progenitor of soybean, *G. soja* or wild soybean, is the most easily accessible to breeders and has a wealth of diversity preserved in the USDA soybean collection (Carter et al., 2004). Thus, it may be an excellent source of new agronomic genes and traits (Lee et al., 2008). Wild soybean has the same chromosome number as the cultivated soybean, crosses freely *via* insect or manual hybridization, and progeny are usually completely fertile (Singh and Hymowitz, 1988; Weber, 1950).

The wild soybean has been introduced into North American breeding programs and used as a source of small seed to develop natto soyfoods cultivars (LeRoy et al., 1991; Johnson et al., 2001; Cui et al., 2004). Nevertheless, *Glycine soja* has not been exploited as a

source of yield alleles by commercial soybean breeders. However, the species appears to have the potential to improve soybean yield (Concibido et al., 2003; Li et al., 2008). Two putative yield alleles from *G. soja* accession plant introduction (PI) 245331 were identified by Li and Pfeiffer (2008) in a population developed through an advanced backcrossing strategy (Tanksley and Nelson, 1996). The alleles were associated with a 6.3% increase in seed yield. Concibido et al. (2003) discovered a *G. soja* allele from PI 407305 associated with a 9.3% increase in seed yield in a recurrent *G. max* parent. However, when the allele was introgressed into six higher yielding elite soybean cultivars, the positive effect of the *G. soja* allele did not appear to hold up. Using inbred lines derived from a cross of *G. max* 'N7103' (Carter, et al. 2003) X *G. soja* PI 366122 (USDA-ARS, NGRP, 2013), Delheimer (2012) identified three genomic regions from *G. soja* which appeared to have a positive effect on seed yield. Several breeding lines developed from the biparental *G. max* x *G. soja* hybridization carried these three genomic regions and yielded as well as the *G. max* parent over multiple locations.

Although these results are encouraging signs that soybean breeders may find useful yield alleles in wild soybean, more and clearer examples are needed to conclusively make the case that wild soybean genome is a valuable source of alleles for seed yield in applied breeding programs. An alternative method to detect yield alleles in wild soybean is to employ heterosis (hybrid vigor) as signal of their presence. Burton and Brownie (2006) suggested that heterosis for yield might be used in *G. max* to detect parental combinations that are most likely to segregate for yield alleles, and thus, produce the most desirable inbred offspring. We suggest that this concept may be extended to wild soybean and used to detect

positive yield alleles in that species as well. Because it is difficult to produce enough F_1 hybrid seed by manual cross-pollination to perform standard yield tests, researchers have turned to F_2 seed and successfully identified heterosis in soybean (Burton and Brownie, 2006).

The wild soybean has prostrate growth habit and F_1 and F_2 plants derived from *G. max* x *G. soja* hybridizations are typically not agronomic. Williams (1948), Weber (1950), and Delheimer (2012) reported that the performance of F_1 and F_2 plants from these crosses as undesirable and resembling the wild soybean parent more than the adapted parent. Examination of heterosis in bi-parental *G. max* x *G. soja* crosses is, thus, not practical. Despite the poor adaptability of segregating *G. max* x *G. soja* populations in general, we are fortunate that adapted inbred lines have been developed which derive 50% of their pedigree from wild soybean and are progeny of a bi-parental cross of *G. max* x *G. soja* (Delheimer, 2012). Hybrids derived by crossing such adapted lines back to the original *G. max* parent would produce agronomic progeny that could be evaluated easily for heterosis. Because such hybrids trace their pedigree to only one unique *G. max* source those exhibiting high-parent heterosis or substantial midparent heterosis for yield would present strong evidence that yield alleles exist in the only other parental source, wild soybean.

The objectives of this study were to 1) examine heterosis in 19 F_2 bulks derived from the cross of *G. max* parental cultivar N7103 and breeding lines which were themselves developed from a cross between N7103 and *Glycine soja* PI 366122, 2) assess whether heterosis could detect the presence of yield alleles derived from wild soybean, and 3) ascertain the possibility of predicting heterosis using genetic markers and agronomic traits.

MATERIALS AND METHODS

Parental Germplasm and Population Development

This study was conducted using recombinant *Glycine max* x *Glycine soja* inbred lines that were developed by Jacob Delheimer (2012) and Dr. Thomas Carter (personal com., 2012). The inbred lines were developed from a single cross between *G. max* 'N7103'and *G. soja* accession PI 366122. The *G. max* cultivar N7103 (PI 615695) is a maturity group (MG) VII small-seeded natto-type that was developed for potential use in the Japanese soyfoods export market. This cultivar has good yield and excellent lodging resistance. N7103 was derived from the mating of NTCPR90-143 x 'Pearl' (Carter et al., 1995). Pearl was derived from 'Vance' x G80-1515, where Vance has the pedigree 'Essex' (Smith and Camper, 1973) x unknown soybean, and G80-1515 has the pedigree 'Pickett 71' (Hartwig, et al. 1971) x 'Bedford' (Hartwig and Epps, 1978). The breeding line NTCPR90-143 originated from the cross 'Gasoy 17' (Baker and Harris, 1979) x Vance, where Gasoy 17 was the product of selection from 'Bragg' (Hinson and Hartwig, 1964) x 'Hood' (Johnson, H.W., 1960; USDA-ARS GRIN Database).

The MG IV *G. soja* accession PI 366122 originated from Japan and has an indeterminate main stem, prostrate growth, a hard seed coat, and small seed (1.7 g per 100 seed; USDA-ARS, NGRP, 2012). The PI 366122 was selected for use as a parent in Delheimer's (2012) research because it exhibited healthy leaf color and a reduced level of bacterial pustule (*Xanthomonas campestris pv. glycines*) in comparison to other wild soybean accessions grown in NC.

The complete mating history of the development of the breeding lines from the cross of these two parents was described by Delheimer (2012). Briefly, the cross between N7103 and PI 366122 was made in 2002 and repeated in 2003, creating two replicate populations. The F_2 populations were bulk harvested and followed by pedigree selection for *G. max*-like appearance until the F_4, at which point, F_4 derived progeny rows were produced. From the selection efforts in the two populations, a total of over 200 upright breeding lines were developed and tested agronomically. For the present study, 18 lines of similar maturity to the *G. max* parent were selected. An additional *G. max* x *G. soja* breeding line, G07-6012, was also selected for study. This line was developed through pedigree selection by Dr. Roger Boerma (personal com., 2012) at the University of Georgia, has the same pedigree as the other breeding lines, and was developed from the same F_1 hybrids made in 2002 for selection in NC.

G. max x *G. soja* Recombinant Inbred Lines Chosen for Study

The 19 lines used for this study were selected based on yield performance, shattering resistance, and a maturity date that was roughly the same as the original *G. max* parent, N7103. Furthermore, the 19 lines were chosen in an effort to capture a diverse spectrum of phenotypic characteristics in the *G. max* x *G. soja* inbred lines, such as leaf shape and seed coat color (Appendix A).

Genotypic information on each breeding line was obtained by Delheimer (2012). The 1,536 Universal Soybean Linkage Panel 1.0 (Hyten et al., 2010) was employed to assay single nucleotide polymorphic (SNP) markers for N7103, PI 366122, and the 19 breeding lines using the GoldenGate assay and analyzed on the Illumina BeadStation 500G (Illumina,

San Diego, CA) at Dr. Cregan's USDA-ARS lab in Beltsville, MD (Hyten et al., 2008). Allele calling for each locus was accomplished with the GenomeStudio software (Illumina, San Diego, CA). The marker data from Delheimer (2012) showed that the percent of *G. soja* alleles in the 19 breeding lines range from 17% to 40% with a mean of 30% *G. soja* alleles based on 558 polymorphic SNP markers.

Patterns of diversity found within the breeding lines were analyzed by calculating genetic distance using simple matching coefficients (SMC; Sokal and Michener 1958). The SMC were based on 558 single nucleotide polymorphism (SNP) markers using the DISTANCE procedure in SAS (SAS Institute, 2007). In the present study, the resulting genetic distance matrix (GD_M) was truncated to include only those lines to be crossed with the *G.* max parent and subjected to multidimensional scaling (MDS) analysis, using the MDS procedure in SAS. The parents were more heavily weighted in the analysis than the original *G. max* and *G. soja* parents to highlight the relation of the breeding lines to the parents, in the manner described by Delheimer (2012).

Hybrid Seed Production

During the summer of 2011, the 19 breeding lines and N7103 were planted at the Clayton research station in NC. Each of the 19 lines (♀) was backcrossed with N7103 (♂; the original *G. max* parent of the breeding lines). Over 150 crossing attempts were made with each line. The crossing efforts produced over 800 putative F_1 seed. About 42 F_1 seeds from each cross plus seed of the parental types were sent to the USDA-ARS Tropical Agriculture Research Station winter nursery in Isabella, Puerto Rico and planted in December 2011 for seed increase. The F_1 seed were planted at seven seed per ten foot (3 m) of row (seed every

20 inches [50.4 cm]) in a lighted field where the lights were turned on at night for six hours for six weeks to extend day length, thereby to delay flowering and increase yield. As a quality control strategy, tissue samples were taken from each F_1 plant and a selected panel of twelve SNP molecular markers were assayed by DuPont Pioneer to verify that each plant was, indeed, a true cross. Approximately ten plants were determined to be other than true crosses and were removed. Each F_1 plant was harvested individually with most plants yielding over 700 F_2 seed. Upon return of the seed from the winter nursery, all seed were screened visually and very few contaminants were identified.

Yield Trials

Sufficient F_2 seed were produced from the F_1 plants to perform standard yield trials at the following four locations in 2012: the Lower Coastal Plain Tobacco Research Station Caswell Farm at Kinston, NC, the Tidewater Research Station at Plymouth, NC, the University of Georgia Plant Science Farm near Athens, GA, and the DuPont Pioneer station near Kinston, NC (Figure 2.1). A nested split-plot experimental design was employed to compare F_2 bulks to the parental performance. Each of the 19 whole plots was randomly assigned to a family (two parents and their F_2 bulk). The F_2 bulk and their two parents (N7103 and a *G. max* X *G. soja* derived breeding line) from a family were randomly assigned to the three subplots within each whole plot. Three replications were employed at each location. The plots were planted in three-rows at the Caswell and Plymouth locations. The GA and DuPont Pioneer locations used four-row yield performance plots. For the F_2 bulks, each three or four-row plot was planted with F_2 seed tracing to only one F_1 plant. This procedure was used so that F_2 segregation ratios could be observed. All F_2 bulks segregated

appropriately in the field indicating that the putative F_2 seed came from true F_1 hybrids. The plots were 17 feet long (5.8 m) at the Caswell and Plymouth locations with 10 seeds planted per foot (0.3 m) with 38 inch (0.97 m) row spacing. The Caswell location was planted in a Lumbee sandy loam (Typic Endoaquult) on June 6, 2012 and the Plymouth location was planted in a Portsmouth loam (Typic Umbraquult) on May 22, 2012. The GA location was planted on June 6, 2012 in a Altavista loamy coarse sand (Aquic Hapludult) at 9-10 seeds per foot (0.3 m) in 20 feet-long rows (6.1 m) with 30 inch (0.76) row spacing. The plots at the DuPont Pioneer location were planted in 36 inch (0.92 m) bedded rows, 15 feet long (4.6 m) with 10 seeds planted per foot (0.3 m) on a Torhunta loam (Typic Humaquepts) on June 6, 2012. Standard soybean production practices were applied to all test plots. Notes on flower color and pubescence color were taken to confirm the identity of each plot. Before harvest, the rows were end trimmed to create a uniform row length at each location. The plots were trimmed to a length of 15 feet (4.6 m) and the center row was harvested for seed yield at the Caswell and Plymouth locations. The GA location was end trimmed to twelve feet (3.7 m) and the two middle rows were harvested in the four-row plots. The plots at DuPont Pioneer location were end-trimmed to eleven feet (3.4 m) and the two middle rows were harvested in the four-row plots. Plant maturity was recorded as the first day on which approximately 95% of the pods reached maturity. After maturity, plant height was recorded as the average of three random plants within the center row of a plot, and lodging was visually scored using a scale of 1 (erect) to 5 (prostrate). After harvest, 100-seed weight and yield per plot was determined at an 8% moisture level. All plot yields were converted to kilograms per hectare for analysis.

Statistical Analysis

Data generated from field trials were subjected to statistical analysis, initially, using the GLM procedure in SAS to generate residuals and identify potential outliers. Cook's D was also calculated in SAS to help identify potential outliers. Data points were removed from the dataset if their associated residuals exceeded three times the standard deviation of the square root of the mean square error and if their Cook's D value was large (SAS Institute Inc, 2004). In total, yield data were deleted for four plots as a result of post-hoc analyses, and data from three yield plots were lost because of poor stands after planting.

Subsequently, the data were analyzed using the MIXED procedure of SAS (SAS Institute Inc, 2004). Analysis of variance (ANOVA) was performed for each location and then over locations for the following traits: yield, plant height, maturity, 100-seed weight, and lodging. Family and genotype within family were treated as fixed effects; all other effects were treated as random. Least square means were estimated using the LSMEANS statement in SAS. Pearson correlation coefficients of genotypic means to agronomic performance of traits and heterosis values were computed using the PROC CORR statement of SAS (SAS Institute Inc, 2004).

Midparent heterosis (MPH) and high-parent heterosis (HPH) were defined as:

$$MPH = F_2 - MP$$

$$HPH = F_2 - HP$$

where: MP = midparent value

HP = high-parent value

Nineteen pre-planned single degree-of-freedom orthogonal contrasts were employed to test the significance of MPH using a one-tailed F-test (i.e. test the hypothesis that each F_2 bulk was not better than the parental average). Pre-planned orthogonal contrasts were also employed to test the significance of the parental comparisons, *G. max* x *G. soja* breeding line parent *versus* the *G. max* N7103 adapted parent. HPH for yield was calculated by comparing each F_2 bulk population with the mean yield value of the *G. max* parent N7103. For all other traits, the HPH value of the F_2 bulk populations was calculated by comparing each F_2 bulk population with its respective HP mean. For all HPH comparisons, a least significant difference (LSD) value was calculated using a one-tailed test and employing the genotype by location mean square as an estimate of error. Heterogeneity of the genotype by location mean squares was tested among families (David, 1952) and they were pooled if they did not significantly ($P \leq 0.05$) differ from each other.

In addition, we estimated the degree of heterosis that might be expected in a F_2 population when heterosis is derived only from positive alleles from the *G. max* parent, assuming complete dominance, and no overdominance or epistasis. The maximum possible yield expected (MYE) from the adapted *G. max* parental alleles in the F_2 population was defined as ¾ of the *G. max* parent yield + ¼ of the inbred *G. max* x *G. soja* breeding line yield. Yield of each F_2 bulk was compared against the appropriate MYE value (F_2 – MYE) and a one-tailed LSD was employed to test the hypothesis that F_2 bulks were not better than the MYE.

The sign test was also employed to determine the probability of observing positive MPH, given the chance of observing positive or negative MPH is equally likely ($P = 0.5$).

Relation between SNP markers in Breeding Line Parents and Heterosis in Progeny

The relation between F_2 heterosis in the bulk progeny and the percent of SNP alleles from *G. soja* in the parental breeding line (derived from the original *G. max* x *G. soja* hybridization) was determined for each trait by regression and by employing Pearson correlation coefficients computed in SAS (SAS Institute Inc, 2004). To ascertain the impact of individual SNP markers on heterosis, multiple linear regression (MLR) analysis (forward and stepwise methods; Rawlings et al., 1998) was performed on the mean of dependent variables MPH and HPH for each of the 19 F_2 populations, with independent variables consisting of 23 SNP markers identified by Delheimer (2012) as having significant ($P \leq 0.05$) positive allelic effects on yield derived from the *G. soja* parent. Delheimer (2012) identified the 23 SNPs based on single factor (individual one degree of freedom) analyses and by culling linked markers which also produced similar positive effects. The stepwise selection models were used to identify genomic regions from the 23 markers that might influence MPH and HPH. Approximately 7% of the marker calls in the data set were missing at random. To facilitate statistical analysis, missing marker data were imputed by employing the frequency of the alleles at each locus to estimate the missing value. For example, if 14 of the 19 genotypes exhibited the *G. soja* SNP allele at a particular marker (value 1), while three genotypes had the *G. max* allele (value 0) and two genotypes were missing data, then the missing data would be replaced with 14/17, or approximately 0.8. The REG procedure in SAS was used to conduct MLR analysis (SAS Institute Inc, 2004). A significance of 0.40 was used as the level for a SNP to enter the model, while a significance level of 0.15 was used for a SNP to remain in the model.

The stepwise selection models were also used to identify genomic regions across all 558 polymorphic loci that might influence MPH. Missing data across the 558 markers were imputed in the same method as described previously. The markers identified as significant ($P \leq 0.15$) in the initial MLR analysis were subject to forward and stepwise regression again to verify the markers were significant ($P < 0.05$) in the model. Lastly, the GLM procedure was used to verify the model was significant under Type I sums of squares. Final parameter estimates of individual marker effect on MPH in the F_2 bulk progeny were obtained from the GLM procedure. The estimates were doubled to reflect expected effects in F_1 progeny. The effect of the markers on *G. soja* x *G. max* breeding line yield was also obtained with the GLM procedure. To visualize *G. soja* gene action, the estimate of effect on MPH in F_1 progeny, *G. soja* x *G. max* breeding lines, and *G. max* parents (zero effect from *G. soja*) was plotted for each marker.

RESULTS

Maturity

The mean maturity of the F_2 bulks overall was slightly later (+2 days) than the mean maturity of the parental lines. However, this small discrepancy may have been an artifact of the protocol used for maturity scoring which introduced a slight bias toward late maturity for the F_2 bulk means. Maturity is recorded as the first date at which 95% of the pods are brown. Many of the F_2 bulks segregated for maturity in the field within plots (three to five days), which made it difficult to record the exact date of maturity of F_2 bulks using the definition above. Other studies also report difficulties in scoring maturity in a heterogeneous population (Weiss et al., 1947; Leffel and Hanson, 1961). Thus, in general, the accuracy of

maturity data was likely biased slightly toward later maturity in F_2 bulks in comparison to the parental lines. The MPH for maturity ranged from -3 to +2 days and HPH ranged from -5 to 2 days. In eleven F_2 bulks, maturity for MPH was significantly ($P \leq 0.05$) and slightly later maturing than the MP (Table 2.1).

Heterosis for Yield in F_2 Bulks

The ANOVA indicated the presence of significant ($P \leq 0.01$) variation among the F_2 bulks and among parents for all traits. The genotype by location interaction was also significant ($P \leq 0.01$) for every trait except maturity. However, the genotype by location interactions, when present were typically smaller than the main genotypic effects (Table 2.2).

In general, the best performing inbred genotype in the test was the adapted *G. max* parent, N7103. The only exception was the *G. max* x *G. soja* derived breeding line, NMS4-1-65, which has been shown in other studies to be high yielding (Dr. Thomas Carter, personal com., 2013; Table 2.1). For MPH of seed yield, the genotype by location interaction and parental genotype by location interactions were very similar in magnitude. Thus, a pooled genotype by location error term was used in F-tests to identify MPH for yield. Overall, the F_2 bulks yielded about 3% greater (not significant, $P \leq 0.05$) than the mean MP value and about 4% less than the mean HP (not significant, $P \leq 0.05$). Fifteen of the 19 F_2 bulks yielded numerically higher than the MP value and two were significantly ($P \leq 0.05$) higher yielding: NMS4-44-329 x N7103 and NMS-112-594 x N7103 (Table 2.1). The MPH of these two combinations was consistently positive at each of the locations and significantly ($P \leq 0.05$) positive at two of the four locations (Table 2.3). The mean percent MPH exhibited in these crosses was +9% and +10%, respectively, over locations. The MPH of these F_2 bulks

was not influenced by genotype by location interaction, however, a significant ($P \leq 0.01$) interaction with location in the parental lines of NMS-112-594 x N7103 was observed (Table 2.2). For HPH, two F_2 bulks yielded higher numerically, but not significantly ($P \leq 0.05$) better than HP *G. max* N7103 over locations (NMS4-44-329 x N7103 and NMS4-175-709 x N7103; Table 2.1). Notably, NMS4-44-329 x N7103 was consistently positive for HPH at each location and significant ($P \leq 0.05$) at one location (Table 2.3).

Across all 19 families, one F_2 bulk from the cross of NMS4-44-329 and N7103 significantly ($P \leq 0.05$) outperformed the respective MYE (FAM 4; Table 2.1). Although the yield of N7103 in that family was the lowest across the entire experiment, the difference between the F_2 bulk yield and the MYE was also significant at two of the four locations (Table 2.3).

Positive MPH was observed in 15 of the 19 families. The sign test indicated that the chance of observing either 15 or more families with positive MPH, or 4 or fewer families with negative MPH in 19 families in a two-tailed test is $P = 0.02$.

Heterosis for Traits Other than Seed Yield

For all traits other than seed yield (100-seed weight, height, and lodging), the mean MPH genotype by location interaction was significantly ($P \leq 0.05$) different from the average parental genotype by location interaction. Therefore, these two genotype by location interaction terms were not pooled in calculating error terms for F-tests of MPH over locations (Table 2.2). The mean MPH of families for 100-seed weight was small and ranged from -0.2 to 0.5 g/100-seed over locations. The HPH for 100-seed weight was typically negative and ranged from -1.6 to 0.4 g/100-seed. The MPH for 100-seed weight in six F_2 bulks was

positive and significant ($P \leq 0.05$), but small as it ranged from 0.3 g/100 seed to 0.5 g/100 seed (Table 2.4).

The F_2 plant height averaged over families was significantly ($P \leq 0.01$) taller than the mean MP value (increase of 5 cm) and F_2 height was significantly ($P \leq 0.05$) greater than the MP for six of the F_2 bulks. The MPH ranged from -5 to 20 cm among bulks. The HPH for height ranged from -4 to 4 cm. Only one combination, NMS5-17-3-22 x N7103, was observed to be significantly ($P \leq 0.05$) taller than the HP for plant height (4 cm; Table 2.5).

In general, lodging was less in each F_2 cross combination than for the *G. max* x *G. soja* breeding line parents, but none of the F_2 bulks lodged less than the *G. max* adapted parent N7103. Over all the genotypes, the mean lodging score for the F_2 bulks (2.5) was intermediate to the mean lodging scores of the *G. max* x *G. soja* breeding lines and the N7103 *G. max* adapted parent (3.0 and 1.9, respectively). Only one F_2 bulk from the cross of NMS5-70-6-129 x N7103 had a significantly ($P \leq 0.05$) greater lodging score in comparison to the MP, while most of the F_2 bulks had significantly ($P \leq 0.05$) greater lodging scores from the HP (

Table 2.6).

Association of Agronomic Traits with Heterosis for Yield

Yield increases in soybean are sometimes accompanied by later plant maturity, taller plant height, and increased lodging. However, only two families showed significant ($P \leq 0.05$) MPH for yield (NMS4-44-329 x N7103 and NMS-112-594 x N7103) and this heterosis could not be readily explained by agronomic traits. Both of these F_2 bulk populations were slightly later in maturity than the MP value (~2 days later), likely as a result of the slight bias

in recording of maturity data for the F_2 bulks. The 100-seed weight and lodging scores for these heterotic F_2 bulks were intermediate to the parents (Tables 2.4 and 2.6). The cross of NMS4-44-329 x N7103 was taller in height than both parents, whereas the cross of NMS-112-594 x N7103 was intermediate to the parents (Table 2.5).

The 19 families exhibited a range in MPH, thus prompting an examination of the correlation of agronomic traits with MPH across families (Table 2.7). On a family mean basis, MPH for yield was significantly ($P \leq 0.05$) associated with maturity and lodging (r = 0.55 and r = 0.65, respectively). MPH for yield did not have a significant ($P \leq 0.05$) association with 100-seed weight, which indicates that the yield increase observed in the F_2 bulks was most likely due to a greater number of seeds. We also calculated correlations after deleting four families, which exhibited negative MPH for yield. In this data set, increased lodging was the only trait that associated significantly ($P \leq 0.05$) and positively with MPH for yield (r = 0.62). These results suggest that increased lodging is likely the only trait in this study to be associated with positive MPH for yield (Table 2.7 and 2.8).

Genetic Analysis of Parents

The MDS analysis revealed a three-dimensional representation of the genetic distance between the 19 *G. max* X *G. soja* inbred breeding lines and their parents (Figure 2.3). The GD_Ms similarity coefficient estimates for the 19 breeding lines and two parents ranged from 0.55 (between NMS5-12-1-9 and N7103) to 0.82 (between NMS5-268-1-568 and N7103; Table 2.9).

The relation between the percent *G. soja* alleles calculated in the 19 breeding lines and the yield performance of each breeding line across all four locations indicated there was

a negative linear trend (r = -0.37), though not significant ($P \leq 0.1$) between yield of the inbred breeding line and its percentage of *G. soja* alleles (Figure 2.4; Table 2.7). In previous yield trials of the breeding lines conducted by Delheimer (2012) and Dr. Thomas Carter (personal com., 2013), the performance of the 19 breeding lines in correlation with percent *G. soja* alleles supported the negative trend and found it significant ($P \leq 0.05$; r = -0.67; 2008-2012 yield data across three or more locations, data not shown).

Relation of Genotypic Markers to Yield Heterosis

A positive and significant ($P \leq 0.05$) correlation was found between the percent of *G. soja* SNP alleles in the breeding line parent and MPH for yield (r =0.60; Table 2.7; Figure 2.5 and 2.6). Across all families, the yield performance of the *G. max* x *G. soja* breeding line yield was also negatively and significantly ($P \leq 0.05$) associated with MPH for yield (r = -0.50). These results, taken together, indicated that in general the lower yielding breeding lines which had a higher percentage of SNP alleles from *G. soja* contributed most to MPH for yield.

The percentage of *G. soja* alleles in the breeding line parent also had positive and significant ($P \leq 0.05$) association with 100-seed weight (r = 0.51) and MPH for 100-seed weight (r = 0.55; Figure 2.7 and 2.8). No significant ($P \leq 0.05$) associations were found between other traits and percent *G. soja* SNP alleles. The HPH for yield was not significantly ($P \leq 0.05$) associated with percent *G. soja* SNP alleles.

The forward method of regression identified five markers that explained more than 45% of the variation in MPH for yield. Three of these five markers were also found in stepwise regression as markers that explain over 40% of the MPH for yield (Table 2.10).

One marker selected (BARC-035255-07160) was previously identified by Delheimer (2012) through MLR as one of three markers significantly ($P \leq 0.05$) contributing to yield from the *G. soja* parent. Delheimer (2012) reported that this marker positively affected yield, whereas the results indicated in the current study show that the marker causes a negative effect on yield. The five markers identified in the current study were also identified as being significant in single marker analysis in Delheimer (2012) for effect on yield, as the percent of check (BARC-035255-07160, BARC-027552-06609, BARC-040651-07808, and BARC-044373-08692) and seed oil content (BARC-022387-04319).

The comparison of the three potential yield genes identified by Delheimer (2012) through MLR in the derived *G. max* x *G. soja* inbred lines to the F_2 bulk MPH produced unclear results in the current study. The *G. max* x *G. soja* breeding line with all three potential yield genes was NMS4-1-65 (Table 2.11). This line was the highest yielding breeding line in this study and also out-yielded the *G. max* N7103 adapted parent. However, the MPH for yield observed in the cross between NMS4-1-65 and N7103 was -158 kg/ha (-5%). Although the negative MPH was not significant, -158 kg/ha was the lowest mean MPH value observed across all locations. The negative MPH observed in this cross was also consistent at each location, except the Caswell, NC location (Table 2.3).

Out of 558 polymorphic markers, a set of 15 SNP markers from the *G. soja* parent were found to significantly ($P \leq 0.05$) effect MPH (Table 2.12). The marker effect on MPH showed both positive and negative estimates; six markers showed a positive effect from *G. soja* and nine markers showed a negative effect from *G. soja*. Numerically, these results suggest *G. max* alleles do not mask all *G. soja* allele effects in the F_2 bulk progeny. The 15

markers that significantly ($P \leq 0.05$) influenced MPH also showed a positive and negative impact on yield in the *G. max* x *G. soja* breeding lines. The plot of effects from *G. soja* on F_1 progeny and the parental line showed gene action that modeled that of overdominance, under-dominance, complete dominance, and partial dominance (Figure 2.9). The figure indicated that some *G. soja* alleles cause a positive heterotic effect in the hybrid progeny as well as have a positive effect in inbred lines.

DISCUSSION

Heterosis for Yield

Heterosis has been often observed in soybean. Many published estimates of heterosis in soybean were based on only a few spaced F_1 plants, because it was very difficult and very time-consuming to obtain large quantities of hybrid seed for standard yield plot testing (Palmer, et al. 2001). However, other studies have successfully reported heterosis for yield in soybean employing replicated standard yield plots in the F_1 and F_2 generation across multiple environments. In standard plots of F_1 yield testing Brim and Cockerham (1961) found a mean of +20% HPH and +28% MPH in two cross combinations. Nelson and Bernard (1984) reported +3% HPH and +8% MPH in 27 crosses using male sterility to facilitate F_1 hybrid development. Lewers et al. (1998) reported a mean of +5% MPH in 36 crosses. Burton and Brownie (2006) found a mean of +11% HPH and +13% MPH in two crosses. In standard yield plots of F_2 bulks Gizlice et al. (1993) found a mean of +9% MPH in ten crosses, Manjarrez-Sandoval et al. (1997) found a mean of +3% HPH and +7% MPH in 24 crosses, and Burton and Brownie (2006) reported a mean of +7% HPH and +9% MPH

in two cross combinations. These reports indicate that heterosis for seed yield is usually positive and fairly common in soybean.

Unlike the current study, all of the previous studies examined heterosis in crosses involving only *G. max* pedigree. Heterosis for yield has not been examined in *G. max* x *G. soja* hybrids using standard yield plots, primarily because F_1s, bulk F_2 and F_3 populations, and early backcross generations are undesirable and resemble the wild soybean more than the cultivated form in those generations (Williams, 1948; Weber, 1950; Delheimer, 2012). The current study is the first to our knowledge to report heterosis for yield in the F_2 generation involving wild soybean pedigree and using standard yield trial plots.

Use of Heterosis to Detect Yield Alleles in Wild Soybean

Heterosis has been interpreted to reflect gene action other than additive. The gene action behind heterosis is generally based on the hypotheses of overdominance, dominance, and/or epistasis (Crow, 2000). The theory behind using heterosis to test for yield genes is dependent upon the latter two hypotheses, dominance and epistasis. The underlying concept is that both heterosis for yield and yield *per se* of inbred progeny are generally regarded as complexly inherited and controlled by many alleles with small effects. Alleles that contribute to yield in a heterotic state should also contribute to yield under inbred conditions for gene actions other than overdominance. Dominant and epistatic interactions can be fixed in inbred line development. Thus, heterosis could be a predictor of inbred line performance (Brownie and Burton, 2006). It was our intent to use 'heterosis for yield' as a signal of the presence of yield alleles in the wild soybean. Our theory is that although yield alleles which function in an inbred state in soybean, would not necessarily show heterosis for yield, those

that do would be sufficiently numerous and/or have sufficiently large effects to be detected in heterotic yield responses. Marker analysis of the current study indicated that both positive and negative heterosis impact yield in the F_2 bulk progeny.

Given the history of parental materials and mating structure for the present study significant ($P \leq 0.05$), positive HPH would have clearly indicated that *G. soja* alleles contributed to the yield performance of F_2 bulks. The substantial MPH for seed yield which we detected, also supported by the sign test, indicated that it is likely that the wild soybean accession PI 366122 carries yield alleles, but the results were somewhat less conclusive. Inter-genotypic competition (blend response) has been reported to cause increases in yield of mixed populations and could have influenced the heterosis we observed in the F_2 progeny (Gizlice et al., 1989a; Gizlice et al., 1989b). While most scenarios for producing MPH for yield in our F_2 populations reflect an underlying basis that wild soybean does indeed provide yield alleles, a less likely alternate scenario exists in which MPH could be generated for seed yield, based only on the occurrence of positive heterotic yield alleles derived from the *G. max* parent. In this 'worst case scenario' argument against the presence of yield alleles derived from wild soybean, the dominant alleles from the *G. max* parent would simply mask the less favorable alleles which might reside in the wild soybean, assuming that heterosis is derived primarily from dominance and not overdominance or epistasis. For example, using a two allele locus model, the F_2 bulk segregates in a 1:2:1 ratio (e.g. ¼AA, ½ Aa, ¼aa), assuming that 'A' allele is derived from *G. max* that 'a' is derived from *G. soja*. In a scenario of all positive yield alleles being completely dominant and derived from the *G. max* parent, N7103, ¾ of the plants in the F_2 bulk will express the dominant yield phenotype of

N7103 (¼ AA + ½ Aa) and ¼ of the plants in the F_2 bulk will express the yield phenotype derived from *G. soja* via the breeding line parent (¼ aa). Thus, extending this concept to multiple loci and employing the same assumptions, the maximum yield heterosis expected (MYE) for yield of an F_2 bulk is expected to be ¾ the yield of *G. max* N7103, plus ¼ the yield of the *G. max* x *G. soja* breeding line parent. Any F_2 bulks yielding greater than the MYE would be evidence that yield alleles from wild soybean have a positive impact on yield. In this study, one F_2 bulk (NMS4-44-329 x N7103) yielded significantly ($P \leq 0.05$) greater than the MYE mean estimate. This cross combination also had significant ($P \leq 0.05$) MPH (253 kg/ha, 9%), was numerically higher than the HP for seed yield, and was consistently positive for MPH and HPH at all locations. These results are strong evidence that alleles from wild soybean PI 366122 contributed to yield increases in the F_2.

Analyses of marker effect on MPH did not seem to support the hypothesis that all *G. max* alleles mask *G. soja* alleles when the alleles are in the heterozygous state. The estimates show evidence that some alleles cause a positive heterotic effect where as other alleles cause a negative heterotic effect. Some of the positive effects from the *G. soja* allele were greater than the effect of the marker in the breeding line. Breeding lines with evidence of positive effects from the *G. soja* parent may be of particular interest for future use in a breeding program as it may be a source of yield potential.

Predicting Heterosis in Populations Containing Wild Soybean Pedigree

In general, agronomic traits were not predictive of MPH for yield. Only lodging appeared to have a significant ($P \leq 0.05$) and stable association with MPH for yield. In contrast, we detected a strong and positive association ($r = 0.60$) between the MPH for yield

and the percent of *G. soja* alleles present in inbred *G. soja* derived breeding lines. In previous soybean studies, genetic distance (as measured by relatively few markers) was reported as a poor predictor for yield heterosis (Manjarrez-Sandoval et al., 1997, Cerna et al., 1997). However, the positive and significant ($P \leq 0.05$) relation between MPH for yield in the F_2 and percent *G. soja* alleles observed in the current study indicates genetic dissimilarity at the molecular level can be predictive of favorable combinations for yield heterosis. It is important to recognize that the previous studies used older technology such as RFLPs and isozymes to determine genetic distance between lines. With the advent of newer technology such as SNP markers, high-throughput may enhance the power of determining genetic distance and provides researchers with a clearer picture of the genetic distance between individuals.

Implications to Soybean Breeding and Future Considerations

Heterosis may be a useful signal to identify which parental breeding line combinations have the best potential for pure line cultivar development. However, further study is needed to determine if this is true for the *G. max* x *G. soja* breeding lines in the current study's population. One approach is to develop recombinant inbred lines (RILs) from the hybrids that have the greatest and least heterosis. The performance of the RILs could then be used to determine the efficacy of using F_2 heterosis as a tool to prioritize populations for breeding and selection. Any inbred lines higher yielding than the original parents from the high MPH group would indicate that heterosis could be used as a selection index. Ultimately, the evidence of heterosis in the F_1 or F_2 generations under standard yield trial plots could be used as a predictor of inbred line performance. For applied plant breeding,

this can help plant breeders know which populations to advance to maximize rate of genetic gain or yield potential.

Although the evidence of MPH in the current study is strong in two F_2 bulks, these observations could be strengthened and validated by measuring heterosis in F_1 hybrids. Burton and Brownie (2006) reported that inbreeding depression caused yield declines of up to 16% from the F_1 generation to the F_2 generation in soybean. Thus, F_1 heterosis should be greater and therefore easier to detect than F_2 heterosis.

All of the breeding lines in the current study may have many positive yield alleles from wild soybean; however, not all the yield alleles may be detectable by measuring heterosis. In fact, it is possible that only a few yield alleles from *G. soja* may be detectable through heterosis. A follow-up quantitative trait loci (QTL) study would be useful to help identify QTL for yield. The F_2 bulks that had high heterosis in the current study would likely be the best starting populations for QTL analysis.

So far we do not know if the yield alleles, that caused the heterosis we observed in this study, will hold up in other soybean populations. Population specificity of these alleles might be of consideration for future testing and the alleles need to be tested for their durability in a wider array of backgrounds. Since the *G. max* adapted parent for this study, N7103, is small seeded for use as a natto cultivar, it is somewhat lower yielding than conventional cultivars. N7103's lower yielding capacity may have influenced why we saw positive heterosis and it is possible the evidence of yield alleles *via* heterosis might not hold up in other populations. Concibido et al. 2003 found a *G. soja* alleles that caused a 9.3%

yield increase, however, after introgression into six elite soybean cultivars, it did not appear to be durable.

Lastly, the breeding lines crossed in this study were derived from a cross of only one *G. soja* accession PI 366122. It is important to extend the search of agronomic alleles and determine whether heterosis is observed and could ultimately indicate if other *G. soja* accessions in the USDA soybean germplasm collection have useful yield genes that soybean breeders can exploit.

REFERENCES

Baker S.H. and H.B. Harris. 1979. Registration of Gasoy 17 Soybeans1 (Reg. No. 121). Crop Sci. 19:130-130.

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

Burton, J.W., and C. Brownie. 2006. Heterosis and inbreeding depression in two soybean single crosses. Crop Sci. 46:2643-2648.

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

Carter, T.E.J., J.W. Burton, M.R. Villagarcia, Z. Cui, X. Zou, M.O. Foundation, D.T. Bowman, A.S. Niewoehner. 2003. Registration of 'N7103' Soybean. Crop Sci. 43: 1128-1128.

Carter, T.E.J., R.L. Nelson, C.H. Sneller and Z.L. Cui. 2004. Genetic diversity in soybean. p. 303–416. In H.R. Boerma, and J.E. Specht (eds.) Soybeans: Improvement, production, and uses. Agronomy Monographs 3rd ed. No. 16 ed. ASA-CSSA-SSSA, Madison, WI, USA.

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

Concibido, V.C., B. La Vallee, P. Mclaird, N. Pineda, J. Meyer, L. Hummel, J. Yang, K. Wu, and X. Delannay. 2003. Introgression of a quantitative trait locus for yield from *Glycine soja* into commercial soybean cultivars. Theor. Appl. Genet. 106:575-582.

Crow, J.F. 2000. The rise and fall of overdominance. Plant Breeding Reviews. p.225-257.

Cui, Z., A.T. James, S. Miyazaki, R.F. Wilson, and T.E. Carter Jr. 2004. Breeding specialty soybeans for traditional and new soyfoods. In Soybeans as Functional Foods and Ingredients, AOCS Press, Champaign, IL264-322.

David, H.A. 1952. Upper 5 and 1% points of the maximum F-ratio. Biometrika. 39(3-4): 422-424.

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

Delheimer, J.C. 2012. Agronomic and molecular analysis of populations developed from a single cross of adapted x wild soybean. (Under the direction of Dr. Thomas E. Carter, Jr.). Ph.D. dissertation. North Carolina State University.

Gizlice, Z., T.E. Carter, and J.W. Burton. 1993. Genetic diversity in North American soybean. 2. prediction of heterosis in F_2 populations of southern founding stock using genetic similarity measures. Crop Sci. 33:620-626.

Gizlice, Z., T.E. Carter, and J.W. Burton. 1989a. The impact of maturity and genotype on blend performance in group-V and group-VII soybean cultivars. Agron. J. 81:559-562.

Gizlice, Z., T.E. Carter, J.W. Burton, and T.H. Emigh. 1989b. Partitioning of blending ability using 2-way blends and component lines of soybean. Crop Sci. 29:885-889.

Hartwig, E.E. and J.M. Epps. 1978. Registration of Bedford Soybean1 (Reg. No. 118). Crop Sci. 18:915-915.

Hartwig, E.E., J.M. Epps, C.J. Edwards. 1971. Registration of Pickett 71 Soybean1 (Reg. No. 87). Crop Sci. 11:603-603.

Hinson, K. and E.E. Hartwig. 1964. Registration of Bragg and Hardee soybeans. Crop Sci. 4:664–665.

Hyten, D., Q. Song, I. Choi, M. Yoon, J. Specht, L. Matukumalli, R. Nelson, R. Shoemaker, N. Young, and P. Cregan. 2008. High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. Theor Appl Genet 116:945-952.

Hyten, D., I. Choi, Q. Song, J.E. Specht, T.E. Carter, Jr., R.C. Shoemaker, E. Hwang, L.K. Matukumalli, and P.B. Cregan. 2010. A high density integrated genetic linkage map of soybean and the development of a 1536 universal soy linkage panel for quantitative trait locus mapping. Crop Sci 50: 960-968.

Johnson, H.W. 1960. Registration of Soybean varieties, VII. Agronomy Journal. Vol. 52 No. 11 pp. 660 pp.

Johnson, S.L., W.R. Fehr, G.A., Welke, and S.R., Cianzio. 2001. Genetic Variability for Seed Size of Two- and Three-Parent Soybean Populations. Crop Sci. 41: 1029–1033.

Kabelka, E.A., B.W. Diers, W.R. Fehr, A.R. LeRoy, I.C. Baianu, T. You, D.J. Neece, and R.L. Nelson. 2004. Putative alleles for increased yield from soybean plant introductions. Crop Sci. 44:784-791.

Lee, J., J. Yu, Y. Hwang, S. Blake, Y. So, G. Lee, H.T. Nguyen, and J.G. Shannon. 2008. Genetic diversity of wild soybean (*Glycine soja* Sieb. and Zucc.) accessions from South Korea and other countries. Crop Sci. 48:606-616.

Leffel, R.C. and W.D. Hanson. 1961. Early generation testing of diallel crosses of soybean. Crop Science 1: 169–174.

LeRoy, A., W. Fehr, and S. Cianzio. 1991. Introgression of genes for small seed size from *Glycine soja* into *G. max*. Crop Sci. 31:693-697.

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

Li, D., T.W. Pfeiffer, and P.L. Cornelius. 2008. Soybean QTL for yield and yield components associated with *Glycine soja* alleles. Crop Sci. 48:571-581.

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

Ma, J., and R. Nelson. 2012. Evaluation of fertile lines derived from the hybridization of *Glycine max* and *G. tomentella* (2*n*=78) is a wild, perennial relative. M.S. thesis, University of Illinois at Urbana-Champaign.

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

Palmer, R.G., J. Gai, H. Sun and J.W. Burton. 2001. Production and evaluation of hybrid soybean. In Plant Breed. Rev. (ed.). p. 263-307.

Rawlings, J.O., S.G. Pantula, and D.A. Dickey. 1998. Applied regression analysis: a research tool. 2nd Ed. Springer-Verlag, New York, NY.

Singh, R.J., and T. Hymowitz. 1988. The genomic relationship between *Glycine max* (L) Merr. and *Glycine soja* Sieb. and Zucc. as revealed by pachytene chromosome analysis. Theor. Appl. Genet. 76:705-711.

Singh, R.J., K.P. Kollipara, and T. Hymowitz. 1990. Backcross-derived progeny from soybean and *Glycine tomentella* hayata intersubgeneric hybrids. Crop Sci. 30:871-874.

Smith, T.J. and H.M. Camper. 1973. Registration of Essex Soybean1 (Reg. No. 97). Crop Sci. 13:495-495.

Sokal, R.R., and C.D. Michener. 1958. A statistical method for evaluating systematic relationships. Univ. Kans. Sci. Bull. 38:1409–1438.

Tanksley, S.D., and J.C. Nelson. 1996. Advanced backcross QTL analysis: A method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor. Appl. Genet. 92:191-203.

USDA-ARS National Genetic Resources Program (USDA-ARS NGRP). 2013. *Glycine soja* Sieb. And Zucc. PI 366122. Germplasm Resources Information Network (GRIN). [Online database] National Germplasm Resources Laboratory, Beltsville, MD. Available at http://www.ars-grin.gov/cgi-bin/npgs/acc/display.pl?1274061 (verified 04 April. 2013).

Weber, C.R. 1950. Inheritance and interrelation of some agronomic and chemical characters in an interspecific cross in soybeans, *Glycine max* x *G. ussuriensis*. Iowa Agric Expt Sta Res Bull 374:765-816.

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

Williams, L. 1948. Inheritance In A Species Cross In The Soybean. P. 131-132. Abstract In 1948. Genetics, Vol 33.

TABLES

Table 2.1 Mean maturity and mean yield of parents and F_2 hybrids. Midparent (MP) heterosis (MPH), high-parent heterosis (HPH), and mean heterosis across all hybrids in 2012, in replicated tests averaged across four locations in 2012.

		Maturity Date					Yield										
Family	♀ Breeding Line	♂ N7103 *G. max*	MP	F_2	MPH‡	♀ Breeding Line	♂ N7103 *G. max*	MP	MYE¶	F_2	MPH‡	HPH#	F_2 - MYE¶	MPH	HPH	*G. soja* SNP Alleles§
		----------Oct. 1 = 1 ----------				---------- kg ha^{-1}----------								----------%----------		
1	NMS4-24-288	24	25	24	25	0.5	2913	3324	3119	3222	3240	121	-85	19	+4	-3	31
2	NMS4-1-65	24	24	24	25	1.3*	3480	3326	3403	3365	3245	-158	-80	-120	-5	-2	20
3	NMS4-8-146	25	24	25	25	0.7	2793	3404	3098	3251	3046	-52	-358	-205	-2	-11	25
4	NMS4-44-329	28	25	26	28	2.3**	3086	3249	3167	3208	3461	294*	212	253*	+9	+7	37
5	NMS4-112-594	22	24	23	24	1.6**	2691	3367	3029	3198	3339	310*	-28	141	+10	-1	32
6	NMS4-169-706	23	25	24	25	1.0	2906	3563	3234	3399	3369	135	-194	-30	+4	-5	28
7	NMS4-175-709	24	24	24	25	1.8**	3391	3406	3399	3403	3460	61	53	58	+2	+2	20
8	NMS5-188-1-406	21	24	23	24	1.3*	2845	3308	3077	3193	3158	81	-150	-35	+3	-5	38
9	NMS5-11-1-4	24	24	24	26	2.1**	2944	3408	3176	3292	3256	81	-151	-36	+3	-4	40
10	NMS5-12-1-9	24	24	24	25	0.8	3024	3465	3245	3355	3418	173	-48	63	+5	-1	38
11	NMS5-17-3-22	25	24	25	26	1.7**	2467	3364	2916	3140	3097	181	-268	-43	+6	-8	34
12	NMS5-37-10-70	26	24	25	26	1.0	3017	3275	3146	3211	3255	109	-20	45	+3	-1	32
13	NMS5-98-3-192	26	25	25	27	1.4*	3038	3324	3181	3253	3315	134	-9	63	+4	0	36
14	NMS5-102-1-207	27	24	26	29	3.0**	2584	3476	3030	3253	3207	177	-269	-46	+6	-8	34
15	NMS5-70-6-129	26	24	25	28	2.7**	2806	3386	3096	3241	3269	173	-117	28	+6	-4	29
16	NMS5-101-2-203	27	25	26	27	1.3*	2793	3271	3032	3152	3155	123	-116	4	+4	-4	36
17	NMS5-268-1-568	24	25	24	25	0.9	2733	3425	3079	3252	3145	66	-280	-107	+2	-8	17
18	NMS5-62-4-111	27	25	26	23	-3.3**	3061	3437	3249	3343	3177	-72	-260	-166	-2	-8	27
19	G07-6012	27	25	26	23	-2.8**	3098	3594	3346	3470	3263	-83	-331	-207	-3	-9	21
mean		25‡‡	24‡‡	25	26‡‡	1.0**	2930	3388	3159	3273	3256	97†	-132†	-18	+3	-4	30

*P-value ≤ 0.05, used one-sided t-test H_0: μ = 0, H_A: μ > 0 based on LSDs of 229 kg/ha for yield MPH, 265 kg/ha for yield HPH, and 239 kg/ha for yield F_2 – MYE.

**P-value ≤ 0.01, used one-sided t-test H_0: μ = 0, H_A: μ > 0 based on LSDs of 326 kg/ha for yield MPH, 377 kg/ha for yield HPH, and 340 kg/ha for yield F_2 – MYE.

‡MPH = F_2 – MP.

¶ MYE = ¾ ♂ N7103 *G. max* yield + ¼ ♀ Breeding line yield.

#HPH = F_2 – HP.

§Percent alleles in female breeding line parent based on a set of 558 polymorphic SNP markers.

†95% confidence interval for: yield MPH (-176, 371), and yield HPH (-448, 184).

‡‡ Maturity date LSD: 4.2 days, calculated as average across all families in two-sided t-test.

Table 2.2 ANOVA for yield in 2012 across four locations.

Source of Variation	Breakdown of Variation	df‡	Mean Squares§ Yield (kg ha^{-1})	100-Seed Weight (g 100 seeds $^{-1}$)	Maturity (Oct. 1 = 1)	Height (cm)	Lodging (1-5 scale)
LOC		3	32440501**	103.008**	1325.71**	2540.96**	61.981**
REP(LOC)		8	1514715**	3.007**	92.46**	148.31**	1.026**
FAMILY		18	448976**	25.180**	35.27**	157.45**	2.397**
LOC*FAMILY		54	172626	0.615	7.18	34.35**	0.347
FAMILY*REP(LOC)		144	145218**	0.306**	5.39**	14.84**	0.187**
GENOTYPE(FAMILY)		38	907720**	14.552**	21.57**	53.23**	4.325**
	F2 vs PARENTS(FAM)	19	182956	0.709**	11.44**	5.12**	0.214
	F2 vs PARENTS(FAM 1)	1	116513	1.767*	1.68	9.67	0.170
	F2 vs PARENTS(FAM 2)	1	153383	0.533	9.00*	5.82	0.309
	F2 vs PARENTS(FAM 3)	1	21973	0.304	4.01	15.21	0.420
	F2 vs PARENTS(FAM 4)	1	716545*	0.380	38.55**	16.91	0.258
	F2 vs PARENTS(FAM 5)	1	767786*	0.101	20.06**	2.67	0.170
	F2 vs PARENTS(FAM 6)	1	145307	0.414	7.35	35.47	0.056
	F2 vs PARENTS(FAM 7)	1	29651	0.532	25.68**	68.47*	0.003
	F2 vs PARENTS(FAM 8)	1	52439	0.934	12.50*	124.84**	0.222
	F2 vs PARENTS(FAM 9)	1	51899	1.417*	34.72**	2.67	0.056
	F2 vs PARENTS(FAM 10)	1	239444	1.537*	4.50	20.54	0.003
	F2 vs PARENTS(FAM 11)	1	173013	1.529*	23.36**	374.31**	0.003
	F2 vs PARENTS(FAM 12)	1	94208	0.422	8.68	16.04	0.125
	F2 vs PARENTS(FAM 13)	1	160864	0.002	14.27*	4.33	0.093
	F2 vs PARENTS(FAM 14)	1	264572	0.819	66.94**	99.77**	0.354
	F2 vs PARENTS(FAM 15)	1	238975	1.299*	58.68**	60.03*	0.781*
	F2 vs PARENTS(FAM 16)	1	120707	1.253*	12.50*	40.67	0.347
	F2 vs PARENTS(FAM 17)	1	34775	0.060	6.12	55.23*	0.056

*P-value $\leq$ 0.05.

**P-value $\leq$ 0.01.

‡Some missing values at some locations.

§F2 vs PARENTS(FAM) for yield tested the probability that F_2 is better than MP, the equivalent of a one-sided t-test.

Table 2.2 cont.

Source of Variation	Breakdown of Variation	df‡	Mean Squares: Yield (kg ha^{-1})	100-Seed Weight (g 100 seeds $^{-1}$)	Maturity (Oct. 1 = 1)	Height (cm)	Lodging (1-5 scale)
	F2 vs PARENTS(FAM 18)	1	41777	0.159	86.68**	16.90	0.281
	F2 vs PARENTS(FAM 19)	1	52329	0.001	62.28**	11.74	0.354
	PARENT vs PARENT(FAM)	19	1625777**	28.411**	17.23**	10.08	8.453**
	PARENT1 vs PARENT2(FAM 1)	1	1014221*	0.107	2.04	114.08*	6.510**
	PARENT1 vs PARENT2(FAM 2)	1	143853	64.518**	1.04	0.01	0.667
	PARENT1 vs PARENT2(FAM 3)	1	2238295**	4.524**	2.04	36.30	3.010**
	PARENT1 vs PARENT2(FAM 4)	1	112605	39.300**	42.75**	1.74	14.301**
	PARENT1 vs PARENT2(FAM 5)	1	2746353**	7.107**	13.50	66.01	19.260**
	PARENT1 vs PARENT2(FAM 6)	1	2590525**	1.233	15.04	5.21	2.667**
	PARENT1 vs PARENT2(FAM 7)	1	1347	65.770**	0.38	21.68	2.344**
	PARENT1 vs PARENT2(FAM 8)	1	1285131**	42.613**	42.67**	158.70**	7.042**
	PARENT1 vs PARENT2(FAM 9)	1	1289536**	25.585**	0.17	66.01	0.167
	PARENT1 vs PARENT2(FAM 10)	1	1167556**	63.375**	0.17	1.20	11.344**
	PARENT1 vs PARENT2(FAM 11)	1	4833809**	45.871**	18.38*	291.41**	16.667**
	PARENT1 vs PARENT2(FAM 12)	1	397279	45.568**	40.04**	8.53	1.042*
	PARENT1 vs PARENT2(FAM 13)	1	396540	51.226**	7.26	16.67	4.688**
	PARENT1 vs PARENT2(FAM 14)	1	4224506**	0.585	45.34**	50.90	18.648**
	PARENT1 vs PARENT2(FAM 15)	1	2020558**	7.718**	18.38*	46.88	10.010**
	PARENT1 vs PARENT2(FAM 16)	1	1366982**	20.350**	20.17*	52.01	15.042**
	PARENT1 vs PARENT2(FAM 17)	1	2866711**	20.498**	1.04	23.41	12.042**
	PARENT1 vs PARENT2(FAM 18)	1	849795*	10.507**	35.04**	93.63*	14.260**
	PARENT1 vs PARENT2(FAM 19)	1	1344156**	23.354**	21.94*	1.88	0.905
LOC*GENOTYPE(FAMILY)		114	153013**	0.398**	3.48	13.71**	0.208**
	F2 vs PARENTS*LOC(FAM)	57	151618**	0.285**	2.29	10.08	0.157**

*P-value $\leq$ 0.05.

**P-value $\leq$ 0.01.

‡Some missing values at some locations.

§F2 vs PARENTS(FAM) for yield tested the probability that F_2 is better than MP, the equivalent of a one-sided t-test.

Table 2.2 cont.

Source of Variation	Breakdown of Variation	df‡	Mean Squares				
			Yield (kg ha^{-1})	100-Seed Weight (g 100 seeds $^{-1}$)	Maturity (Oct. 1 = 1)	Height (cm)	Lodging (1-5 scale)
	F2 vs PARENTS*LOC(FAM 1)	3	102507	0.243	4.79	2.41	0.253*
	F2 vs PARENTS*LOC(FAM 2)	3	101264	0.212	2.74	0.72	0.040
	F2 vs PARENTS*LOC(FAM 3)	3	71734	0.059	1.68	3.20	0.068
	F2 vs PARENTS*LOC(FAM 4)	3	167224	0.267	2.39	3.07	0.158
	F2 vs PARENTS*LOC(FAM 5)	3	52192	0.032	0.35	6.87	0.096
	F2 vs PARENTS*LOC(FAM 6)	3	186984	0.614*	4.72	2.91	0.176
	F2 vs PARENTS*LOC(FAM 7)	3	101577	0.038	0.46	40.14**	0.096
	F2 vs PARENTS*LOC(FAM 8)	3	45189	0.176	0.54	38.56**	0.139
	F2 vs PARENTS*LOC(FAM 9)	3	61272	0.238	2.02	1.19	0.102
	F2 vs PARENTS*LOC(FAM 10)	3	221464*	1.318**	0.31	5.16	0.670**
	F2 vs PARENTS*LOC(FAM 11)	3	153167	0.592*	1.72	12.91	0.162
	F2 vs PARENTS*LOC(FAM 12)	3	153390	0.161	2.27	8.49	0.023
	F2 vs PARENTS*LOC(FAM 13)	3	83167	0.321	0.94	10.60	0.161
	F2 vs PARENTS*LOC(FAM 14)	3	490335**	0.151	3.62	24.13*	0.022
	F2 vs PARENTS*LOC(FAM 15)	3	522179**	0.273	2.27	7.64	0.022
	F2 vs PARENTS*LOC(FAM 16)	3	114829	0.119	3.50	4.65	0.412**
	F2 vs PARENTS*LOC(FAM 17)	3	120194	0.132	3.38	12.39	0.093
	F2 vs PARENTS*LOC(FAM 18)	3	8418	0.420	4.90	5.88	0.152
	F2 vs PARENTS*LOC(FAM 19)	3	123649	0.045	0.84	0.61	0.174
PARENT vs PARENT*LOC(FAM)		57	154408**	0.511**	4.68*	17.31**	0.261**
	PARENT1 vs PARENT2*LOC(FAM 1)	3	18060	0.113	7.82	24.18*	0.205
	PARENT1 vs PARENT2*LOC(FAM 2)	3	4581	0.502*	5.26	1.29	0.111
	PARENT1 vs PARENT2*LOC(FAM 3)	3	366143**	0.196	7.49	7.15	0.066
	PARENT1 vs PARENT2*LOC(FAM 4)	3	244253*	0.158	2.60	3.81	0.035

*P-value $\leq$ 0.05.

**P-value $\leq$ 0.01.

‡Some missing values at some locations.

Table 2.2 cont.

Source of Variation	Breakdown of Variation	df‡	Mean Squares				
			Yield (kg ha^{-1})	100-Seed Weight (g 100 seeds $^{-1}$)	Maturity (Oct. 1 = 1)	Height (cm)	Lodging (1-5 scale)
	PARENT1 vs PARENT2*LOC(FAM 5)	3	237639*	0.233	2.83	12.78	1.038**
	PARENT1 vs PARENT2*LOC(FAM 6)	3	59833	0.357	2.04	8.36	0.194
	PARENT1 vs PARENT2*LOC(FAM 7)	3	14036	0.856**	2.04	29.76*	0.510**
	PARENT1 vs PARENT2*LOC(FAM 8)	3	40973	0.725*	1.89	3.28	0.014
	PARENT1 vs PARENT2*LOC(FAM 9)	3	75455	0.387	9.17*	3.11	0.028
	PARENT1 vs PARENT2*LOC(FAM 10)	3	140558	2.041**	3.39	19.18	0.122
	PARENT1 vs PARENT2*LOC(FAM 11)	3	849064**	1.158**	2.49	15.87	0.750**
	PARENT1 vs PARENT2*LOC(FAM 12)	3	51197	0.700*	2.15	4.88	0.292*
	PARENT1 vs PARENT2*LOC(FAM 13)	3	165678	0.277	15.36**	10.40	0.350**
	PARENT1 vs PARENT2*LOC(FAM 14)	3	89494	0.705*	4.18	18.29	0.515**
	PARENT1 vs PARENT2*LOC(FAM 15)	3	61721	0.145	2.93	2.29	0.066
	PARENT1 vs PARENT2*LOC(FAM 16)	3	271622*	0.493	3.61	100.07**	0.125
	PARENT1 vs PARENT2*LOC(FAM 17)	3	115508	0.049	7.04	38.29**	0.153
	PARENT1 vs PARENT2*LOC(FAM 18)	3	78742	0.516*	0.82	23.54*	0.038
	PARENT1 vs PARENT2*LOC(FAM 19)	3	49201	0.090	5.82	2.38	0.346*
Residual		297	72867	0.190	3.24	8.52	0.093

*P-value $\leq$ 0.05.

**P-value $\leq$ 0.01.

‡Some missing values at some locations.

Table 2.3 Midparent heterosis (MPH) and high-parent heterosis (HPH) for seed yield across all locations (All) and from each location, Plymouth, NC (PLY); DuPont Pioneer, Kinston, NC (KIN); Athens, GA (GA); and Kinston, NC (CAS) in yield test TCLP 593 in 2012.

			Locations														
			MPH[#]					HPH[#]					F_2 - MYE				
Family	♀ Breeding Line Parent	*G. soja* SNP Alleles (%)[§]	All	PLY	KIN	GA	CAS	All	PLY	KIN	GA	CAS	All	PLY	KIN	GA	CAS
1	NMS4-24-288	31	121	185	67	-155	386*	-85	13	-209	-308	165	19	99	-71	-232	276
2	NMS4-1-65	20	-158	-43	-449	-186	75	-80	-5	-348	-103	163	-120	-24	-399	-145	119
3	NMS4-8-146	25	-52	-268	26	-135	167	-358	-354	-252	-791	-34	-205	-311	-113	-463	67
4	NMS4-44-329	37	294*	85	588**	494*	22	212	176	539*	124	36	253*	130	563**	309*	29
5	NMS4-112-594	32	310*	306*	415*	435*	83	-28	177	203	-32	-461	141	242	309*	201	-189
6	NMS4-169-706	28	135	432**	-269	297	79	-194	154	-598	-171	-160	-30	293	-433	63	-40
7	NMS4-175-709	20	61	269	-251	170	56	53	324	-245	116	19	58	297	-248	143	37
8	NMS5-188-1-406	38	81	69	-100	89	267	-150	-286	-305	-89	78	-35	-109	-203	0	173
9	NMS5-11-1-4	40	81	-22	292	-98	151	-151	-354	27	-358	80	-36	-188	160	-228	115
10	NMS5-12-1-9	38	173	137	-104	650**	9	-48	73	-531	485*	-218	63	105	-318	568**	-104
11	NMS5-17-3-22	34	181	424*	227	-337	281	-268	334	-373	-1244	83	-43	379*	-73	-790	182
12	NMS5-37-10-70	32	109	471**	-5	-182	151	-20	330	-259	-258	107	45	400*	-132	-220	129
13	NMS5-98-3-192	36	134	422*	-87	255	26	-9	481*	-250	-110	-64	63	451*	-168	72	-19
14	NMS5-102-1-207	34	177	-49	916**	-53	-83	-269	-331	316	-521	-497	-46	-190	616**	-287	-290
15	NMS5-70-6-129	29	173	877**	-251	-152	217	-117	661**	-446	-492	-192	28	769**	-348	-322	13
16	NMS5-101-2-203	36	123	371*	54	-179	245	-116	198	-481	-393	213	4	285	-214	-286	229
17	NMS5-268-1-568	17	66	-74	347*	-194	185	-280	-250	14	-552	-331	-107	-162	181	-373	-73
18	NMS5-62-4-111	27	-72	-5	-113	-141	-30	-260	-83	-324	-479	-155	-166	-44	-219	-310	-92
19	G07-6012	21	-83	64	-72	-434	126	-331	-186	-259	-805	-45	-207	-61	-166	-620	40
mean		30	97	192	65	8	127	-132	56	-199	-315	-64	-18	99	-71	-232	276

*P-value ≤ 0.05.

**P-value ≤ 0.01.

[§]Percent alleles in female breeding line parent based on a set of 558 polymorphic SNP markers.

[#]MPH and HPH for yield tested that F_2 is better than MP and F_2 is better than HP, respectively, the equivalent of a one-sided t-test.

Table 2.4 Mean 100-seed weight (SDWT) of parents and F_2 hybrids. Midparent (MP) heterosis (MPH), high-Parent (HP) heterosis (HPH), and mean heterosis across all hybrids in replicated tests averaged across four locations in 2012.

Family	♀ Breeding Line Parent	♀ Breeding Line Parent	♂ N7103 *G. max*, HP	MP	F_2	MPH[‡]	HPH[#]	*G. soja* SNP Alleles (%)[§]
		----------SDWT (grams 100 seeds $^{-1}$)----------						
1	NMS4-24-288	8.3	8.2	8.2	8.7	0.5*	0.4	31
2	NMS4-1-65	11.4	8.1	9.8	10.0	0.2	-1.4**	20
3	NMS4-8-146	9.2	8.3	8.8	8.6	-0.2	-0.6**	25
4	NMS4-44-329	11.0	8.3	9.6	9.9	0.2	-1.1**	37
5	NMS4-112-594	7.3	8.4	7.8	7.9	0.1	-0.4	32
6	NMS4-169-706	8.9	8.4	8.6	8.9	0.2	0.0	28
7	NMS4-175-709	11.6	8.3	9.9	10.2	0.3	-1.4**	20
8	NMS5-188-1-406	11.0	8.3	9.6	10.0	0.3*	-1.0**	38
9	NMS5-11-1-4	10.6	8.5	9.5	9.9	0.4*	-0.6**	40
10	NMS5-12-1-9	11.5	8.3	9.9	10.3	0.4*	-1.2**	38
11	NMS5-17-3-22	11.0	8.3	9.7	10.1	0.5*	-0.9**	34
12	NMS5-37-10-70	11.0	8.3	9.6	9.9	0.2	-1.1**	32
13	NMS5-98-3-192	11.5	8.4	9.9	9.9	0.0	-1.6**	36
14	NMS5-102-1-207	8.7	8.4	8.6	8.9	0.3*	0.2	34
15	NMS5-70-6-129	9.4	8.3	8.9	9.3	0.4*	-0.2	29
16	NMS5-101-2-203	10.0	8.2	9.1	9.5	0.4*	-0.5**	36
17	NMS5-268-1-568	6.6	8.4	7.5	7.4	-0.1	-1.0**	17
18	NMS5-62-4-111	9.7	8.4	9.0	9.2	0.1	-0.5**	27
19	G07-6012	6.4	8.4	7.4	7.4	0.0	-1.0**	21
mean		9.7	8.3	9.0	9.3	0.2**[†]	-0.7[†]	30

*P-value ≤ 0.05.

**P-value ≤ 0.01.

[‡]MPH = F_2 – MP.

[#]HPH = F_2 – HP.

[§]Percent alleles in female breeding line parent based on a set of 558 polymorphic SNP markers.

[†]95% confidence interval: 100-seed weight MPH (0.14, 0.32) and 100-seed weight (-0.99, -0.49).

Table 2.5 Mean height of parents and F_2 hybrids. Midparent (MP) heterosis (MPH), high-parent (HP) heterosis (HPH), and mean heterosis across all hybrids in replicated tests averaged across four locations in 2012.

Family	♀ Breeding Line Parent	♀ Breeding Line Parent	♂ N7103 *G. max*	MP	F_2	MPH[‡]	HPH[#]	*G. soja* SNP Alleles (%)[§]
		---------- Height (cm)----------						
1	NMS4-24-288	104	92	98	101	3	-1	31
2	NMS4-1-65	87	86	86	89	2	1	20
3	NMS4-8-146	98	91	95	98	4	0	25
4	NMS4-44-329	95	93	94	99	5	1	37
5	NMS4-112-594	87	96	92	93	2	-1	32
6	NMS4-169-706	86	88	87	93	6	2	28
7	NMS4-175-709	88	82	85	93	9*	2	20
8	NMS5-188-1-406	101	86	94	106	13**	2	38
9	NMS5-11-1-4	78	87	82	84	2	-1	40
10	NMS5-12-1-9	86	87	87	92	5	2	38
11	NMS5-17-3-22	110	90	100	120	20**	4*	34
12	NMS5-37-10-70	82	85	84	87	3	1	32
13	NMS5-98-3-192	80	87	84	85	2	-1	36
14	NMS5-102-1-207	99	88	94	105	11**	2	34
15	NMS5-70-6-129	99	91	95	103	8*	1	29
16	NMS5-101-2-203	98	89	93	100	6	1	36
17	NMS5-268-1-568	95	88	92	98	7*	1	17
18	NMS5-62-4-111	103	90	96	92	-5	-4*	27
19	G07-6012	91	90	90	87	-3	-2	21
mean		93	89	91	96	5**[†]	1[†]	30

*P-value ≤ 0.05.

**P-value ≤ 0.01.

[‡]MPH = F_2 – MP.

[#]HPH = F_2 – HP.

[§]Percent alleles in female breeding line parent based on a set of 558 polymorphic SNP markers.

[†]95% confidence interval: height MPH (2.68, 7.67) and height HPH (-0.31, 1.36).

Table 2.6 Mean lodging of parents and F_2 hybrids. Midparent (MP) heterosis (MPH), high-parent (HP) heterosis (HPH), and mean heterosis across all hybrids in replicated tests averaged across four locations in 2012.

Family	♀ Breeding Line Parent	♀ Breeding Line Parent	♂ N7103 *G. max*	MP	F_2	MPH[‡]	HPH[#]	*G. soja* SNP Alleles (%)[§]
		----------Lodging (1-5 scale)----------						
1	NMS4-24-288	3.0	2.0	2.5	2.6	0.1	0.7**	31
2	NMS4-1-65	2.3	1.9	2.1	2.3	0.2	0.4*	20
3	NMS4-8-146	2.7	2.0	2.4	2.1	-0.2	0.1	25
4	NMS4-44-329	3.5	1.8	2.6	2.8	0.2	1.0**	37
5	NMS4-112-594	3.7	1.9	2.8	2.9	0.1	1.0**	32
6	NMS4-169-706	2.6	1.9	2.3	2.3	0.1	0.4**	28
7	NMS4-175-709	2.5	1.8	2.1	2.2	0.0	0.3*	20
8	NMS5-188-1-406	2.9	1.8	2.3	2.5	0.2	0.7**	38
9	NMS5-11-1-4	2.1	1.9	2.0	1.9	-0.1	0.0	40
10	NMS5-12-1-9	3.1	1.8	2.4	2.4	0.0	0.7**	38
11	NMS5-17-3-22	3.5	1.8	2.6	2.6	0.0	0.8**	34
12	NMS5-37-10-70	2.3	1.9	2.1	2.2	0.1	0.3*	32
13	NMS5-98-3-192	2.8	1.8	2.3	2.4	0.1	0.6**	36
14	NMS5-102-1-207	3.8	1.9	2.9	3.1	0.2	1.2**	34
15	NMS5-70-6-129	3.2	1.9	2.5	2.8	0.3*	1.0**	29
16	NMS5-101-2-203	3.4	1.8	2.6	2.8	0.2	1.0**	36
17	NMS5-268-1-568	3.3	1.8	2.5	2.6	0.1	0.8**	17
18	NMS5-62-4-111	3.4	1.8	2.6	2.4	-0.2	0.6**	27
19	G07-6012	2.5	2.0	2.2	2.0	-0.2	0.0	21
mean		3.0	1.9	2.4	2.5	0.1[†]	0.6**[†]	30

*P-value ≤ 0.05.

**P-value ≤ 0.01.

[‡]MPH = F_2 – MP.

[#]HPH = F_2 – HP.

[§]Percent alleles in female breeding line parent based on a set of 558 polymorphic SNP markers.

[†]95% confidence interval: lodging MPH (0.00, 0.14) and lodging HPH (0.45, 0.77).

Table 2.7 Correlation of genotypic means of MPH and HPH of yield, 100-seed weight, maturity, height, and lodging; and F_2 bulk estimates for all traits, and percent *G. soja* SNP alleles in 2012 TCLP 593 across all locations.

	Yield MPH	Yield HPH	100-Seed Weight MPH	100-Seed Weight HPH	Maturity MPH	Maturity HPH	Height MPH	Height HPH	Lodging MPH	Lodging HPH
Yield MPH	1	0.50*	0.36	0.24	0.59**	0.48*	0.40	0.36	0.49*	0.68**
Yield HPH	.	1	0.27	-0.33	0.41	0.38	-0.05	0.13	0.50*	0.25
100-Seed Weight MPH	.	.	1	0.30*	0.36	0.32	0.46*	0.34	0.41	0.31
100-Seed Weight HPH	.	.	.	1	0.07	0.01	0.06	-0.13	0.16	0.30
Maturity MPH	.	.	.	.	1	0.96**	0.62**	0.69**	0.68**	0.42
Maturity HPH	.	.	.	.	.	1	0.55*	0.64**	0.56*	0.27
Height MPH	.	.	.	.	.	.	1	0.88**	0.37	0.44
Height HPH	.	.	.	.	.	.	.	1	0.44	0.32
Lodging MPH	.	.	.	.	.	.	.	.	1	0.68**
Lodging HPH	.	.	.	.	.	.	.	.	.	1
F_2 Bulk Yield	.	.	.	.	.	.	.	.	.	.
♀ Breeding Line Yield	.	.	.	.	.	.	.	.	.	.
♂ N7103 Yield	.	.	.	.	.	.	.	.	.	.
F_2 Bulk 100-Seed Weight	.	.	.	.	.	.	.	.	.	.
F_2 Bulk Maturity	.	.	.	.	.	.	.	.	.	.
♀ Breeding Line Maturity	.	.	.	.	.	.	.	.	.	.
♂ N7103 Maturity	.	.	.	.	.	.	.	.	.	.
F_2 Bulk Height	.	.	.	.	.	.	.	.	.	.
F_2 Bulk Lodging	.	.	.	.	.	.	.	.	.	.
G. soja SNP Alleles	.	.	.	.	.	.	.	.	.	.

*P-value ≤ 0.05.

**P-value ≤ 0.01.

Table 2.7 cont.

	F_2 Bulk Yield	♀ Breeding Line Yield	♂ N7103 Yield	F_2 Bulk 100-Seed Weight	F_2 Bulk Maturity	♀ Breeding Line Maturity	♂ N7103 Maturity	F_2 Bulk Height	F_2 Bulk Lodging	*G. soja* SNP Alleles (%)
Yield MPH	0.40	-0.50*	-0.27	0.12	0.55*	0.00	-0.15	0.33	0.65**	0.60**
Yield HPH	0.76**	0.45	-0.59**	0.52*	0.37	-0.03	-0.08	-0.23	0.18	0.31
100-Seed Weight MPH	0.16	-0.12	-0.22	0.57**	0.27	-0.08	-0.25	0.37	0.24	0.55*
100-Seed Weight HPH	-0.22	-0.53*	0.24	-0.38	0.10	0.02	0.16	0.33	0.41	0.15
Maturity MPH	0.18	-0.25	-0.41	0.38	0.82**	-0.18	-0.40	0.32	0.41	0.37
Maturity HPH	0.20	-0.15	-0.35	0.37	0.70**	-0.31	-0.41	0.21	0.25	0.25
Height MPH	-0.21	-0.52*	-0.18	0.34	0.46*	-0.17	-0.36	0.83**	0.38	0.23
Height HPH	0.08	-0.24	-0.10	0.36	0.54*	-0.16	-0.27	0.57*	0.26	0.09
Lodging MPH	0.22	-0.09	-0.51*	0.21	0.62**	-0.06	-0.03	0.26	0.68**	0.21
Lodging HPH	0.03	-0.50*	-0.35	0.00	0.51*	0.16	0.09	0.59**	0.98**	0.32
F_2 Bulk Yield	1	0.52*	0.07	0.26	0.14	-0.07	0.04	-0.43	-0.01	0.09
♀ Breeding Line Yield	.	1	-0.05	0.31	-0.25	-0.05	0.12	-0.66**	-0.53*	-0.35
♂ N7103 Yield	.	.	1	-0.48*	-0.39*	-0.03	0.18	-0.18	-0.28	-0.37
F_2 Bulk 100-Seed Weight	.	.	.	1	0.33	0.00	-0.28	0.05	-0.13	0.51*
F_2 Bulk Maturity	.	.	.	.	1	0.41	-0.08	0.29	0.52*	0.38
♀ Breeding Line Maturity	.	.	.	.	.	1	0.34	0.01	0.20	0.08
♂ N7103 Maturity	.	.	.	.	.	.	1	-0.15	0.11	-0.12
F_2 Bulk Height	.	.	.	.	.	.	.	1	0.58**	0.17
F_2 Bulk Lodging	.	.	.	.	.	.	.	.	1	0.26
G. soja SNP Alleles (%)	.	.	.	.	.	.	.	.	.	1

*P-value ≤ 0.05.

**P-value ≤ 0.01.

Table 2.8 Correlation of genotypic means using those F_2 bulks which were numerically positive for yield MPH, against the traits of yield HPH, and MPH and HPH for 100-seed weight, maturity, height, and lodging, F_2 bulk estimates for all traits, and percent *G. soja* SNP alleles in 2012 TCLP 593 across all locations

Variable	Yield MPH	Yield HPH	100-Seed Weight MPH	100-Seed Weight HPH	Maturity MPH	Maturity HPH	Height MPH	Height HPH
Yield MPH	1	0.38	-0.03	0.14	0.34	0.06	-0.09	-0.06
Yield HPH	.	1	-0.12	-0.42	0.02	-0.02	-0.50	-0.28
100-Seed Weight MPH	.	.	1	0.38	0.10	0.11	0.36	0.20
100-Seed Weight HPH	.	.	.	1	0.13	0.05	0.01	-0.16
Maturity MPH	.	.	.	.	1	0.74**	0.25	0.24
Maturity HPH	.	.	.	.	.	1	0.04	0.03
Height MPH	.	.	.	.	.	.	1	0.85**
Height HPH	.	.	.	.	.	.	.	1
Lodging MPH	.	.	.	.	.	.	.	.
Lodging HPH	.	.	.	.	.	.	.	.
♀ Breeding Line Yield	.	.	.	.	.	.	.	.
F_2 Bulk Yield	.	.	.	.	.	.	.	.
F_2 Bulk 100-Seed Weight	.	.	.	.	.	.	.	.
F_2 Bulk Maturity	.	.	.	.	.	.	.	.
F_2 Bulk Height	.	.	.	.	.	.	.	.
F_2 Bulk Lodging	.	.	.	.	.	.	.	.
G. soja SNP Alleles (%)	.	.	.	.	.	.	.	.

**P-value ≤ 0.01.

Table 2.8 cont.

Variable	Lodging MPH	Lodging HPH	F_2 Bulk Yield	♀ Breeding Line Yield	F_2 Bulk 100-Seed Weight	F_2 Bulk Maturity	F_2 Bulk Height	F_2 Bulk Lodging	*G. soja* SNP Alleles (%)
Yield MPH	0.32	0.62*	0.33	0.09	-0.13	0.31	0.16	0.62*	0.33
Yield HPH	0.09	-0.07	0.77**	-0.12	0.37	0.12	-0.43	-0.10	0.15
100-Seed Weight MPH	-0.07	-0.05	-0.15	0.09	0.51	0.03	0.39	-0.06	0.48
100-Seed Weight HPH	0.39	0.28	-0.25	-0.08	-0.49	0.08	0.28	0.41	0.02
Maturity MPH	0.34	0.32	0.06	0.02	0.12	0.75**	0.21	0.35	0.17
Maturity HPH	-0.12	-0.04	0.20	-0.03	0.08	0.42	-0.05	-0.01	-0.08
Height MPH	0.00	0.29	-0.49	0.33	0.21	0.11	0.87**	0.23	-0.05
Height HPH	-0.03	0.22	-0.11	0.05	0.27	0.21	0.63*	0.13	-0.19
Lodging MPH	1	0.70**	-0.14	-0.09	-0.31	0.44	0.29	0.74**	0.00
Lodging HPH	·	1	-0.21	0.23	-0.37	0.41	0.60*	0.98**	0.05
F_2 Bulk Yield	·	·	1	-0.17	0.24	0.03	-0.54*	-0.21	-0.06
♀ Breeding Line Yield	·	·	·	1	0.11	-0.32	0.33	0.18	0.33
F_2 Bulk 100-Seed Weight	·	·	·	·	1	0.15	-0.05	-0.46	0.53*
F_2 Bulk Maturity	·	·	·	·	·	1	0.15	0.43	0.19
F_2 Bulk Height	·	·	·	·	·	·	1	0.57*	-0.01
F_2 Bulk Lodging	·	·	·	·	·	·	·	1	0.02
G. soja SNP Alleles (%)	·	·	·	·	·	·	·	.	1

*P-value ≤ 0.05.

**P-value ≤ 0.01.

Table 2.9 Genetic distance matrix by simple matching coefficients similarity coefficients of the 19 breeding lines, original *G. max* N7103 adapted parent, and original *G. soja* PI 366122 parent on the basis of 558 polymorphic loci.

Genotype	NMS4-24-288	NMS4-1-65	NMS4-8-146	NMS4-44-329	NMS4-112-594	NMS4-169-706	NMS4-175-709	NMS5-188-1-406	NMS5-11-1-4	NMS5-12-1-9
NMS4-24-288	1	.	.	.	.	.	.	.	.	.
NMS4-1-65	0.64	1	.	.	.	.	.	.	.	.
NMS4-8-146	0.88	0.70	1	.	.	.	.	.	.	.
NMS4-44-329	0.56	0.64	0.58	1	.	.	.	.	.	.
NMS4-112-594	0.81	0.61	0.79	0.57	1	.	.	.	.	.
NMS4-169-706	0.90	0.61	0.92	0.52	0.80	1	.	.	.	.
NMS4-175-709	0.65	1.00	0.71	0.66	0.61	0.61	1	.	.	.
NMS5-188-1-406	0.54	0.60	0.50	0.45	0.53	0.51	0.60	1	.	.
NMS5-11-1-4	0.55	0.60	0.53	0.53	0.49	0.48	0.60	0.54	1	.
NMS5-12-1-9	0.52	0.57	0.50	0.49	0.49	0.47	0.57	0.51	0.68	1
NMS5-17-3-22	0.54	0.60	0.53	0.51	0.48	0.52	0.61	0.58	0.61	0.56
NMS5-37-10-70	0.56	0.62	0.57	0.51	0.49	0.51	0.62	0.49	0.54	0.45
NMS5-98-3-192	0.56	0.61	0.55	0.53	0.53	0.55	0.61	0.59	0.61	0.53
NMS5-102-1-207	0.55	0.60	0.53	0.55	0.54	0.52	0.60	0.51	0.51	0.51
NMS5-70-6-129	0.62	0.66	0.58	0.53	0.57	0.54	0.66	0.54	0.52	0.52
NMS5-101-2-203	0.54	0.63	0.54	0.49	0.53	0.49	0.63	0.51	0.50	0.48
NMS5-268-1-568	0.58	0.67	0.58	0.52	0.58	0.59	0.67	0.54	0.53	0.56
NMS5-62-4-111	0.60	0.68	0.59	0.62	0.59	0.56	0.68	0.52	0.57	0.56
G07-6012	0.52	0.64	0.55	0.51	0.56	0.52	0.64	0.51	0.49	0.50
N7103	0.67	0.80	0.69	0.58	0.65	0.67	0.80	0.58	0.57	0.55
PI366122	0.28	0.20	0.19	0.31	0.28	0.23	0.20	0.33	0.37	0.30

Table 2.9 cont.

Genotype	NMS5-17-3-22	NMS5-37-10-70	NMS5-98-3-192	NMS5-102-1-207	NMS5-70-6-129	NMS5-101-2-203	NMS5-268-1-568	NMS5-62-4-111	G07-6012	N7103	PI366122
NMS5-17-3-22	1	.	.	.	.	.	.	.	.	.	.
NMS5-37-10-70	0.54	1	.	.	.	.	.	.	.	.	.
NMS5-98-3-192	0.59	0.51	1	.	.	.	.	.	.	.	.
NMS5-102-1-207	0.50	0.45	0.47	1	.	.	.	.	.	.	.
NMS5-70-6-129	0.53	0.69	0.56	0.52	1	.	.	.	.	.	.
NMS5-101-2-203	0.49	0.70	0.57	0.48	0.71	1	.	.	.	.	.
NMS5-268-1-568	0.54	0.51	0.54	0.64	0.59	0.51	1	.	.	.	.
NMS5-62-4-111	0.60	0.51	0.53	0.75	0.60	0.53	0.67	1	.	.	.
G07-6012	0.50	0.48	0.47	0.54	0.57	0.53	0.71	0.59	1	.	.
N7103	0.61	0.63	0.61	0.62	0.67	0.59	0.82	0.69	0.73	1	.
PI366122	0.30	0.27	0.32	0.31	0.25	0.31	0.15	0.23	0.16	0.00	1

Table 2.10 SNP markers found significantly (P < 0.15) associated with MPH for yield in forward and stepwise regression of 23 SNP markers on MPH for yield and effect of markers on *G. max* x *G. soja* breeding line yield and MPH.

Regression Method[‡]	SNP Marker	Chromosome	Linkage Group	Position (cM)	Associated Traits Found by Delheimer (2012)	Estimate of Marker Effect on *G. max* x *G. soja* Breeding Line[+]	Estimate of Marker Effect on MPH
Forward and Stepwise	BARC-035255-07160	10	O	42	Yield as Percent of Check	96	-129
Forward and Stepwise	BARC-027552-06609	20	I	19	Yield as Percent of Check	21	118
Forward and Stepwise	BARC-022387-04319	8	A2	81	Seed Oil Content	133	130
Forward	BARC-040651-07808	5	A1	3	Yield as Percent of Check	365	72
Forward	BARC-044373-08692	4	C1	81	Yield as Percent of Check	-555	77

[‡]Model R-square for Stepwise = 0.41 and Forward = 0.49

[+]Regression on 23 markers significant (P < 0.15) in modeling only MPH.

Table 2.11 Presence of three potential yield alleles from *G. soja* as reported in Delheimer, 2012, compared to inbred line performance and heterosis results in the present study.

										SNP Marker[†]		
Family	♀ Breeding Line Parent		♂ N7103 *G. max*	MP	MYE¶	F_2	MPH[‡]	HPH[#]	F_2 - MYE¶	BARC-017645-02642	BARC-031311-07043	BARC-035255-07160
			----------Yield (kg ha^{-1})----------									
1	NMS4-24-288	2913	3324	3119	3222	3240	121	-85	19		X	X
2	NMS4-1-65	3480	3326	3403	3365	3245	-158	-80	-120	X	X	X
3	NMS4-8-146	2793	3404	3098	3251	3046	-52	-358	-205		X	H
4	NMS4-44-329	3086	3249	3167	3208	3461	294*	212	253*	X	X	
5	NMS4-112-594	2691	3367	3029	3198	3339	310*	-28	141	H	X	
6	NMS4-169-706	2906	3563	3234	3399	3369	135	-194	-30		X	H
7	NMS4-175-709	3391	3406	3399	3403	3460	61	53	58	X	X	.
8	NMS5-188-1-406	2845	3308	3077	3193	3158	81	-150	-35	X		X
9	NMS5-11-1-4	2944	3408	3176	3292	3256	81	-151	-36			X
10	NMS5-12-1-9	3024	3465	3245	3355	3418	173	-48	63			X
11	NMS5-17-3-22	2467	3364	2916	3140	3097	181	-268	-43			.
12	NMS5-37-10-70	3017	3275	3146	3211	3255	109	-20	45		X	X
13	NMS5-98-3-192	3038	3324	3181	3253	3315	134	-9	63		.	
14	NMS5-102-1-207	2584	3476	3030	3253	3207	177	-269	-46			X
15	NMS5-70-6-129	2806	3386	3096	3241	3269	173	-117	28		X	
16	NMS5-101-2-203	2793	3271	3032	3152	3155	123	-116	4		X	X
17	NMS5-268-1-568	2733	3425	3079	3252	3145	66	-280	-107			
18	NMS5-62-4-111	3061	3437	3249	3343	3177	-72	-260	-166			X
19	G07-6012	3098	3594	3346	3470	3263	-83	-331	-207	X		.

[‡]MPH = F_2 – MP.

¶ MYE = ¾ ♂ N7103 *G. max* yield + ¼ ♀ Breeding line yield.

[#]HPH = F_2 – HP.

[†] 'X' indicates *G. soja* yield allele present at that SNP; 'H' indicates the breeding line was heterozygous at that SNP; "."= missing data.

*P-value ≤ 0.05, used one-sided t-test H_0: μ = 0, H_A: μ > 0 MPH for yield, HPH for yield, and yield F_2 – MYE based on LSDs of229 kg/ha, 265 kg/ha, and 239 kg/ha.

Table 2.12 17 SNP markers found significantly (P < 0.15) associated with MPH from the *G. soja* parent across 558 polymorphic markers and narrowed to 15 SNP markers found significantly (P < 0.05) associated with MPH after subsequent MLR analysis. Marker position and effect on MPH and effect of each marker on *G. max* x *G. soja* breeding line yield.

Regression Method[‡]	SNP Marker	Chromosome	Linkage Group	Position	Significant ($P \leq 0.05$) in Further MLR Analysis	Estimate of Marker Effect on *G. max* x *G. soja* Breeding Line[+]	Estimate of Marker Effect on MPH
				cM		-----Yield (kg ha^{-1})-----	
Forward and Stepwise	BARC-028583-05961	13	F	76	X	-193	114
Forward and Stepwise	BARC-046144-10286	13	F	85	X	117	218
Forward and Stepwise	BARC-035255-07160[+]	10	O	42	X	490	-243
Forward and Stepwise	BARC-011645-00322	16	J	68	X	435	-189
Forward	BARC-038489-10129	8	A2	74	X	-577	121
Forward	BARC-060037-16311	1	D1A	59	X	837	-62
Forward	BARC-041237-07944	13	F	21	X	25	-155
Forward	BARC-029491-06207	10	O	82	X	206	58
Forward	BARC-049601-09082	6	C2	102	X	212	-90
Forward	BARC-043197-08552	18	G	1	X	-195	-19
Forward	BARC-056517-14441	8	A2	135	X	-47	-21
Forward	BARC-048959-10760	14	B2	63	X	295	23
Forward	BARC-018959-03045	15	E	6	X	-233	-26
Forward	BARC-050267-09542	1	D1A	42	X	79	9
Forward	BARC-055571-13451	15	E	82	X	-254	-1
Forward	BARC-008021-00209	10	O	92			0
Forward	BARC-047665-10370	18	G	16			0

[‡]Regression on 558 polymorphic loci significant ($P \leq 0.15$) in modeling only MPH.

[+] Markers not significant ($P \leq 0.15$) for modeling breeding line yield in forward and stepwise regression.

FIGURES

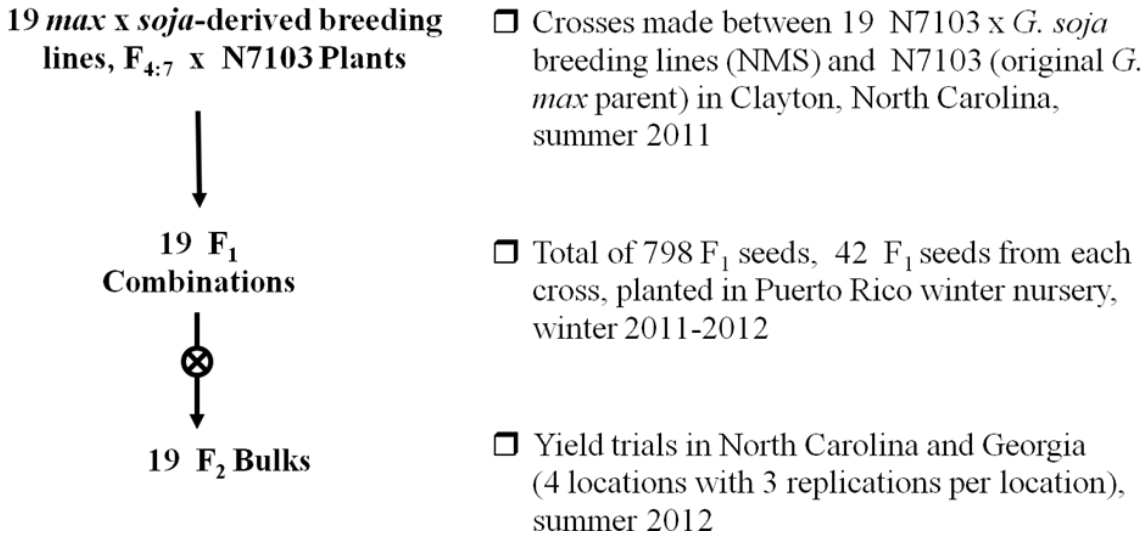


Figure 2.1 Diagram outlining development of F_2 plants and hybrid yield testing.

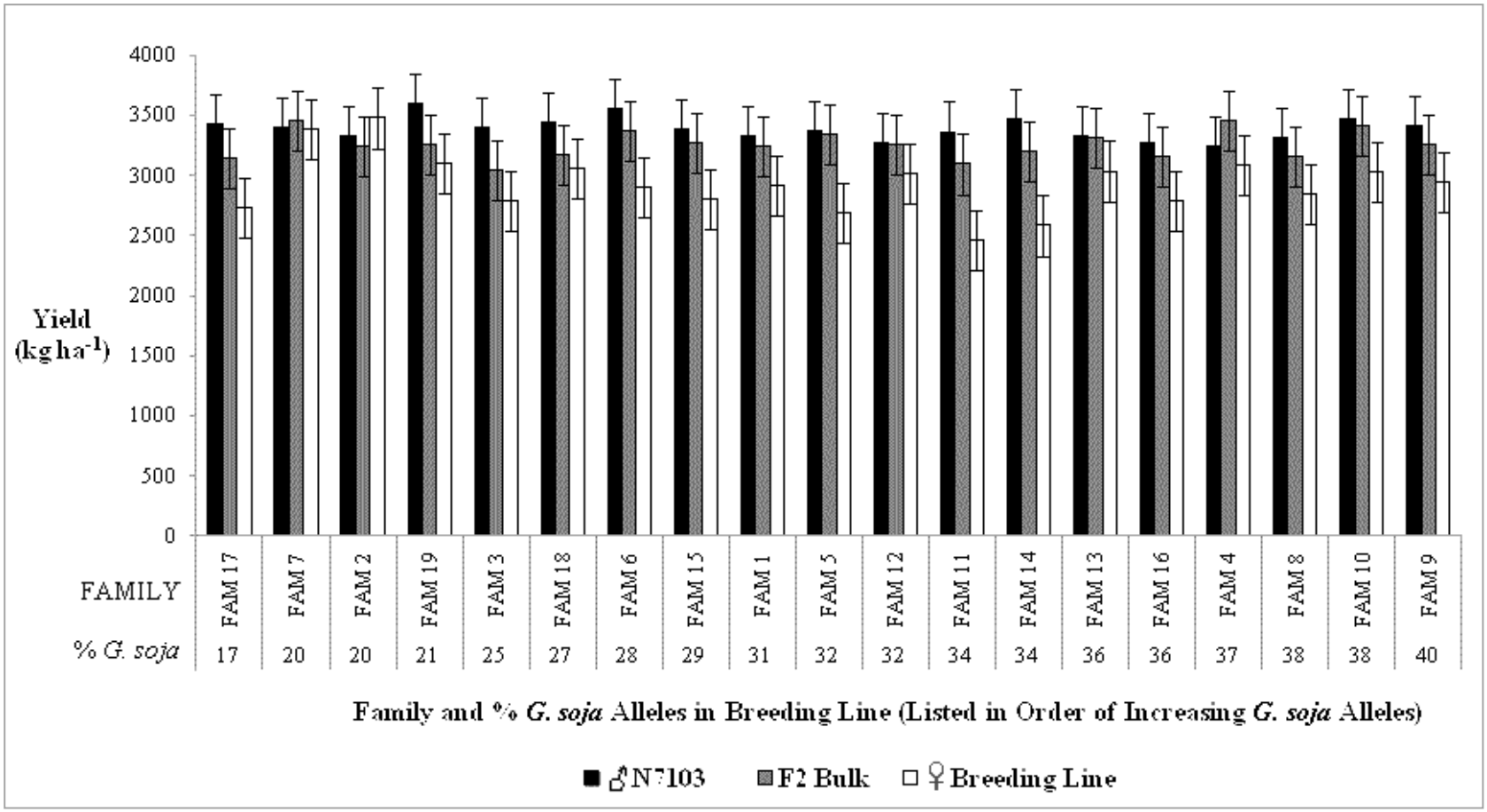
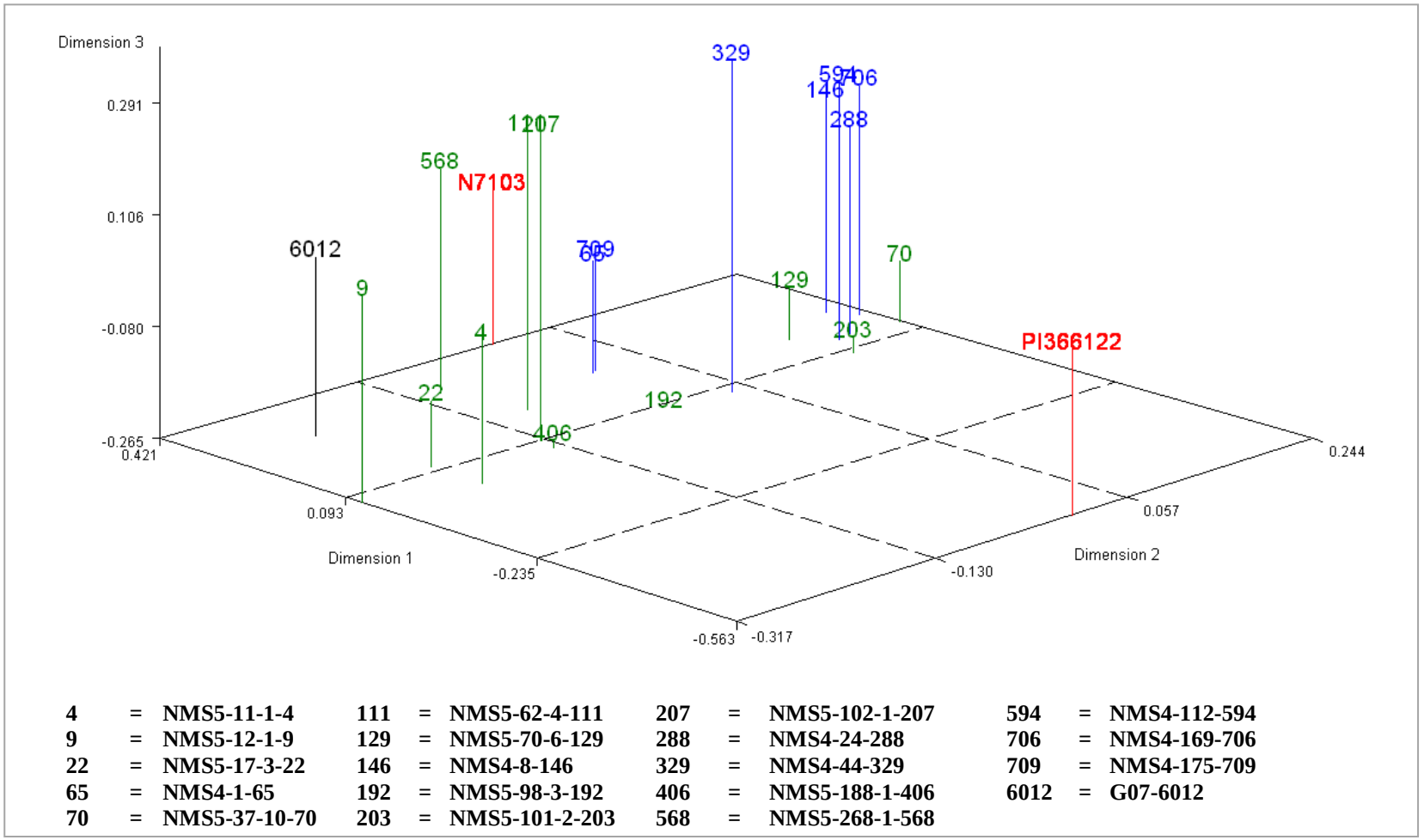


Figure 2.2 Yield performance of parents and F_2 bulks grouped by family and their respective % *G. soja* alleles in breeding line. Percent *G. soja* alleles in female breeding lines based on a set of 558 polymorphic SNP markers.



4	**=**	**NMS5-11-1-4**	**111**	**=**	**NMS5-62-4-111**	**207**	**=**	**NMS5-102-1-207**	**594**	**=**	**NMS4-112-594**
9	**=**	**NMS5-12-1-9**	**129**	**=**	**NMS5-70-6-129**	**288**	**=**	**NMS4-24-288**	**706**	**=**	**NMS4-169-706**
22	**=**	**NMS5-17-3-22**	**146**	**=**	**NMS4-8-146**	**329**	**=**	**NMS4-44-329**	**709**	**=**	**NMS4-175-709**
65	**=**	**NMS4-1-65**	**192**	**=**	**NMS5-98-3-192**	**406**	**=**	**NMS5-188-1-406**	**6012**	**=**	**G07-6012**
70	**=**	**NMS5-37-10-70**	**203**	**=**	**NMS5-101-2-203**	**568**	**=**	**NMS5-268-1-568**			

Figure 2.3 MDS analysis: three dimensional depiction of genetic distance of 18 RILs (blue and green) from Delheimer's (2012) research and one RIL (black) from UGA, derived from the original *G. max* 'N7103' and *G. soja* PI 366122 parents (red). The genotypes in blue were developed in the first breeding cycle. The green colored genotypes were developed in the second breeding cycle.

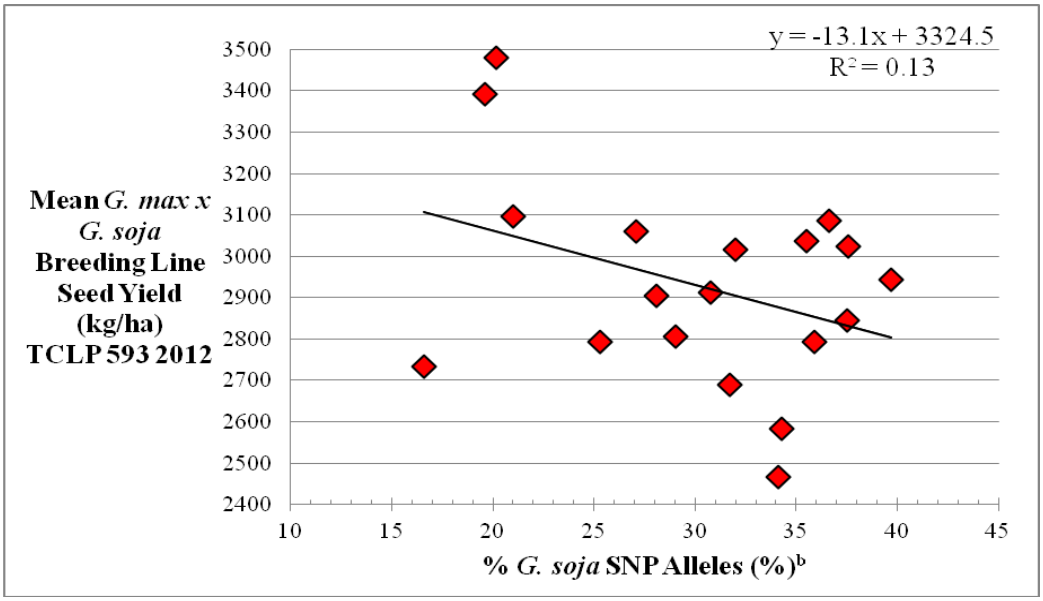


Figure 2.4 The relationship between the percent *G. soja* alleles present in the 19 breeding lines and the yield performance of each line from the 2011 TCLP 593 2012 test with correlation of *r* = -0.37; not significant.

[b]Percent *G. soja* alleles in female breeding lines based on a set of 558 polymorphic SNP markers.

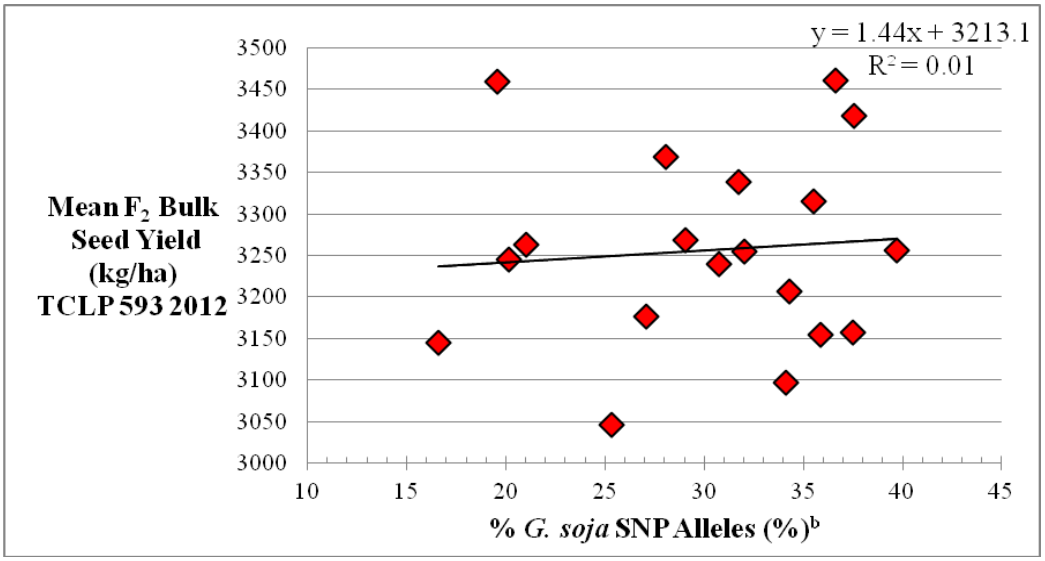


Figure 2.5 Mean seed yield of each F_2 genotype versus percent of *G. soja* alleles in parental lines across all locations in TCLP 593 2012 test with a correlation of *r* = 0.10; not significant.

[b]Percent *G. soja* alleles in female breeding lines based on a set of 558 polymorphic SNP markers.

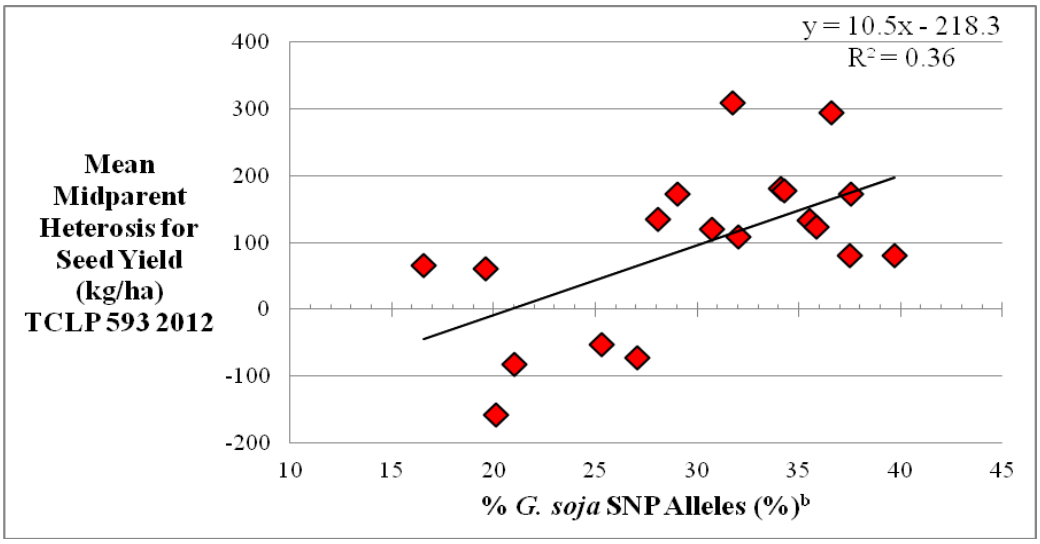


Figure 2.6 Mean percent MPH of each F_2 genotype versus percent of *G. soja* alleles in parental lines across all locations in TCLP 593 2012 with a correlation of *r* = 0.60; significant at P ≤ 0.01.

[b]Percent *G. soja* alleles in female breeding lines based on a set of 558 polymorphic SNP markers.

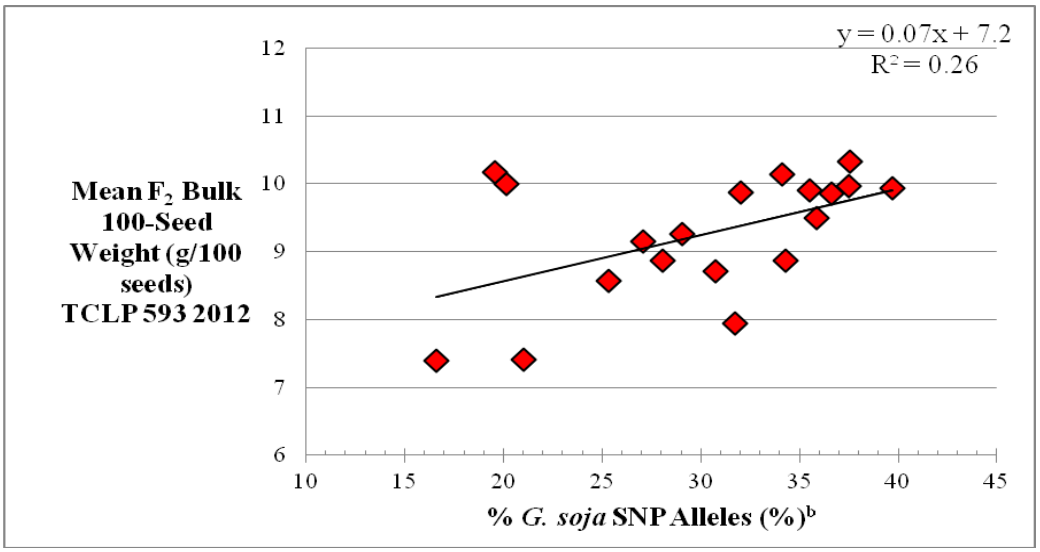


Figure 2.7 Mean F_2 bulk 100-seed weight (g/100 seeds) of each F_2 genotype versus percent of *G. soja* SNP alleles in parental breeding lines across all locations in TCLP 593 2012 with a correlation of *r* = 0.51; significant at P ≤ 0.05.

[b]Percent *G. soja* alleles in female breeding lines based on a set of 558 polymorphic SNP markers.

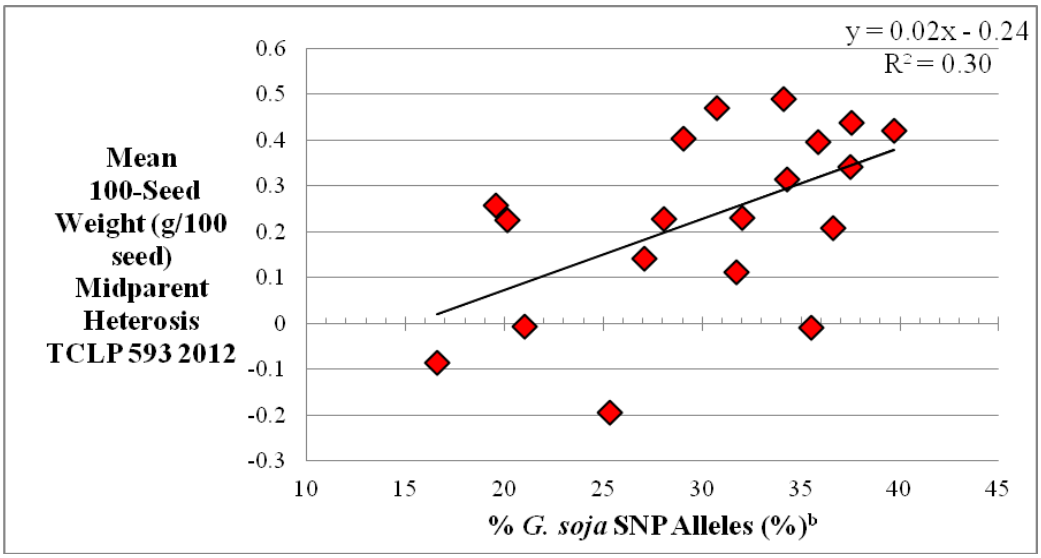


Figure 2.8 Mean 100-seed weight (g/100 seeds) MPH of each F_2 genotype versus percent of *G. soja* SNP alleles in parental breeding lines across all locations in TCLP 593 2012 with a correlation of $r = 0.55$; significant at $P \leq 0.05$.

[b]Percent *G. soja* alleles in female breeding lines based on a set of 558 polymorphic SNP markers.

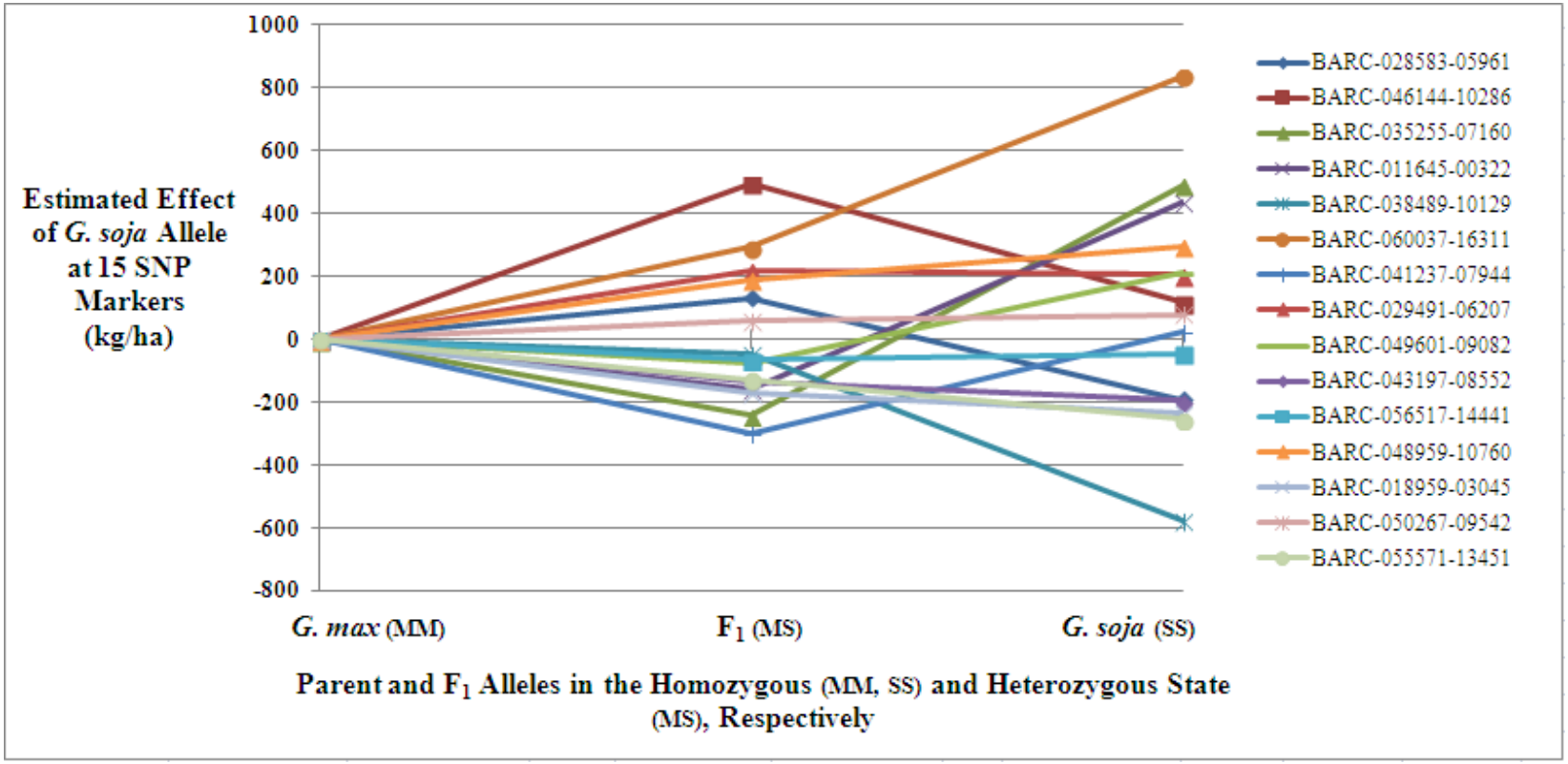


Figure 2.9 Plotted *G. soja* gene action based on the estimate of effect on MPH in F_1 progeny, *G. soja* x *G. max* breeding lines, and *G. max* parent (zero effect from *G. soja*) for markers found significantly ($P \leq 0.05$) associated with MPH in forward and backward regression.

APPENDIX

Appendix A. Agronomic performance and phenotypes of the 19 breeding lines and the checks evaluated and as reported in Delheimer's (2012) research averaged across all locations in TCLP tests.

Genotype	Test Name	Years averaged across	Yield	Flower Color[±]	Pubescence Color[¥]	Leaf Shape[††]	Seed Coat Color[#]	Pod Wall Color[¶]	Plant Height	Lodging[*]	Maturity Group
	TCLP		kg ha^{-1}						cm		
NMS4-24-288	381	2008-2010	2119	W	G	R	Y	L	107	2.5	7.8
NMS4-1-65	383	2008-2010	2694	W	G	N	Y	L	71	1.5	6.3
NMS4-8-146	383	2008-2010	2090	W	G	R	Y	L	94	2.0	7.1
NMS4-44-329	383	2008-2010	2271	P	T	R	BL	S	81	2.0	8.8
NMS4-112-594	383	2008-2010	1901	W	G	R	Y	D	87	3.0	8.1
NMS4-169-706	383	2008-2010	2177	W	G	R	Y	L	92	2.0	7.5
NMS4-175-709	383	2008-2010	2614	W	G	N	Y	L	79	2.0	6.3
NMS5-188-1-406	406A	2009-2010	2045	W	G	R	Y	D	95	2.5	8.9
NMS5-11-1-4	408A	2009-2010	2048	W	T	N	BL	L	70	1.5	9.2
NMS5-12-1-9	408A	2009-2010	1975	W	T	R	Y	D	91	2.5	8.9
NMS5-17-3-22	408A	2009-2010	2070	W	G	R	Y	L	101	3.0	8.4
NMS5-37-10-70	408A	2009-2010	2104	W	G	N	Y/GR	D	75	1.5	8.1
NMS5-98-3-192	409A	2009-2010	2131	W	S	R	Y	D	65	2.0	8.7

*1 indicates all plants erect, 5 indicates all prostrate.

[±] W = White, P=Purple, S=Segregating.

[¥] G = Grey, T = Tawny.

[††] R = Round, N = Narrow.

[¶] L = Light, D = Dark.

[#] Y = Yellow, GR = Green, BL = Black.

Appendix A. cont.

Genotype	TCLP Test	Years averaged across	Yield	Flower Color[±]	Pubescence Color[¥]	Leaf Shape[††]	Seed Coat Color[#]	Pod Wall Color[¶]	Plant Height	Lodging[*]	Maturity Group
			kg ha^{-1}						cm		
NMS5-102-1-207	409A	2009-2010	1936	W	T	R	Y/BL	D	39	3.5	8.4
NMS5-70-6-129	423	2009-2010	1748	W	S	N	GR/BL	D	40	3.0	7.6
NMS5-101-2-203	423	2009-2010	2138	W	G	R	Y	D	41	3.0	8.5
NMS5-268-1-568	423	2009-2010	1748	S	G	R	Y	L	35	2.5	5.8
NMS5-62-4-111	423	2009-2010	1856	W	S	R	BL/BR	L	39	3.0	7.4
G07-6012[§]				W	G	N	Y	L	34	2.0	7.1
N7103[‡†]	408A	2009-2010	2553	W	G	N	Y	L	29	1.5	7.5
NC-Roy[†]	408A	2009-2010	2585	W	G				40	1.5	6.9
NC-Raleigh[†]	408A	2009-2010	2679	W	T				46	1.5	7.5

[§] Line from Dr. Roger Boerma at UGA.
[‡] *G. max* parent.
[†] *G. max* conventional checks.
*1 indicates all plants erect, 5 indicates all prostrate.
[±] W = White, P=Purple, S=Segregating.
[¥] G = Grey, T = Tawny.
[††] R = Round, N = Narrow.
[¶] L = Light, D = Dark.
[#] Y = Yellow, GR = Green, BL = Black.

Appendix B. Agronomic performance and phenotypes of the 19 breeding lines and the checks evaluated in three different environments (NC, LA, and GA) in 2011 and 2012 TCLP 462.

Genotype[+]	Yield	100-Seed Weight	% of Est. Check Yield (N7103)	% *G. soja* SNP Alleles[#]
	kg ha^{-1}	g/100		
NMS4-24-288	2165*	8.6	77	31
NMS4-1-65	2603	12.8	94	20
NMS4-8-146	2387	9.4	83	25
NMS4-44-329	2334	11.6	85	37
NMS4-112-594	2071*	7.7	71	32
NMS4-169-706	2179*	9.3	77	28
NMS4-175-709	2966	12.5	105	20
NMS5-188-1-406	1809*	10.7	66	38
NMS5-11-1-4	1802*	10.3	69	40
NMS5-12-1-9	1876*	10.9	66	38
NMS5-17-3-22	1930*	10.9	71	34
NMS5-37-10-70	2334*	12.0	82	32
NMS5-98-3-192	2199*	11.5	80	36
NMS5-102-1-207	2374*	9.2	83	34
NMS5-70-6-129	2206*	10.0	78	29
NMS5-101-2-203	2132*	10.1	83	36
NMS5-268-1-568	1970*	6.5	71	17
NMS5-62-4-111	2307*	9.3	83	27
G07-6012[§]	2179*	6.4	80	21
N7103‡†	2771	8.7		
NC-Roy†	2556	14.4		
NC-Raleigh†	3221*	14.4		

+ NMS4 types are from the first population and NMS5 types are from the second population.
* Significantly different from the N7103 check based on LSD pair wise comparisons at α 0.05.
# Based on 558 SNP loci.
§ Line from Dr. Roger Boerma at UGA.
‡ *G. max* parent.
†*G. max* conventional checks.

Appendix C. List of the 19 lines crossed to N7103 with their similarity coefficient and their field numbers in Puerto Rico.

Descriptors±	♀ Parent	X	♂ Parent	Descriptors±	F_2 Progeny Name†	Puerto Rico Field Row Number	Number of Putative F_1 Plants in Puerto Rico	Similarity Coefficient§
W G R	NMS4-24-288	X	N7103	W G N	LR-785	265-270	42	0.67
W G N	NMS4-1-65	X	N7103	W G N	LR-786	271-276	37	0.80
W G R	NMS4-8-146	X	N7103	W G N	LR-787	277-282	41	0.69
P T R	NMS4-44-329	X	N7103	W G N	LR-788	283-288	41	0.58
W G R	NMS4-112-594	X	N7103	W G N	LR-789	289-294	41	0.65
W G R	NMS4-169-706	X	N7103	W G N	LR-790	295-300	42	0.67
W G N	NMS4-175-709	X	N7103	W G N	LR-791	301-306	41	0.80
W G R	NMS5-188-1-406	X	N7103	W G N	LR-792	307-312	41	0.58
W T N	NMS5-11-1-4	X	N7103	W G N	LR-793	313-318	30	0.57
W T R	NMS5-12-1-9	X	N7103	W G N	LR-794	319-324	41	0.55
W G R	NMS5-17-3-22	X	N7103	W G N	LR-795	325-330	40	0.61
W G N	NMS5-37-10-70	X	N7103	W G N	LR-797	331-336	31	0.63
W S R	NMS5-98-3-192	X	N7103	W G N	LR-800	337-342	40	0.61
W T R	NMS5-102-1-207	X	N7103	W G N	LR-801	343-348	36	0.62
W S N	NMS5-70-6-129	X	N7103	W G N	LR-802	349-354	38	0.67
W G R	NMS5-101-2-203	X	N7103	W G N	LR-803	355-360	41	0.59
S G R	NMS5-268-1-568	X	N7103	W G N	LR-804	361-366	26	0.82
W S R	NMS5-62-4-111	X	N7103	W G N	LR-805	367-372	37	0.69
W G N	G07-6012	X	N7103	W G N	LR-807	265-270	37	0.73

± W = White, P=Purple, S=Segregating, G = Grey, T = Tawny, R = Round, N = Narrow.

§A measure of genetic distance between two individuals, the breeding line and N7103, based on 558 SNP loci.

†Number after 'LR' indicates row number of breeding line in Clayton, NC crossing block where crosses were made and harvested.

Appendix D. Outliers/extraneous variables taken out of data from test TCLP 593 in 2012 that exceeded the limit of three times the standard deviation of the square root of the mean square error and missing plots.

Entry	Loc	Rep	Family	Code	Plot	Yield	Maturity Date	Lodging	Height	100-Seed Weight	R value	Cook's D	Residual Variation	Standard Deviation Variation
						kg ha^{-1}	Oct 1. = 1		cm	g/100				
Outliers														
N7103	KIN	2	4	11	114	4870	30	2.5	99	9.5	627	0.14	783	253
N7103	PLY	3	19	57	159	1812	24	3	.	7.5	-658	0.17	-945	229
LR-786	GA	1	2	4	130	1086	19	1	66	9.5	-918	0.18	-1175	323
LR-795	GA	1	11	31	135	389	23	2	114	10.6	-850	0.16	-1026	323
Missing Plots														
LR-800	PLY	1	13	37	5	.	.	.	.	.	.	.	.	.
N7103-13	PLY	1	13	38	6	.	.	.	.	.	.	.	.	.
N7103-14	PLY	1	14	41	15	.	.	.	.	.	.	.	.	.

Appendix E. Percent *G. soja* alleles calculated in the inbred breeding lines, F_1 hybrids, and F_2 bulk populations used in this study.

		G. soja alleles[b]			
Family	Hybrid Name	$F_{4:7}$ Breeding Line	F_1	F_2 Chromosomal Positions Heteroyzgous	F_2 Chromosomal Positions Homozygous
		---------- % ----------			
1	LR-785	31	15.4	7.7	3.8
2	LR-786	20	10.1	5.0	2.5
3	LR-787	25	12.7	6.3	3.2
4	LR-788	37	18.3	9.2	4.6
5	LR-789	32	15.9	7.9	4.0
6	LR-790	28	14.0	7.0	3.5
7	LR-791	20	9.8	4.9	2.4
8	LR-792	38	18.8	9.4	4.7
9	LR-793	40	19.9	9.9	5.0
10	LR-794	38	18.8	9.4	4.7
11	LR-795	34	17.0	8.5	4.3
12	LR-797	32	16.0	8.0	4.0
13	LR-800	36	17.8	8.9	4.4
14	LR-801	34	17.1	8.6	4.3
15	LR-802	29	14.5	7.3	3.6
16	LR-803	36	17.9	9.0	4.5
17	LR-804	17	8.3	4.1	2.1
18	LR-805	27	13.5	6.8	3.4
19	LR-807	21	10.5	5.3	2.6
mean		30	15	8	4

[b]Percent alleles in female breeding line parent based on a set of 558 polymorphic SNP markers.

Appendix F. Overall ANOVA for 100-seed weight, maturity, height, and lodging for soybean parents and F_2 hybrids in 2012 Yield Test TCLP 593 across all locations.

		2012				
		Mean Squares				
Source of Variation	df[a]	Yield	100-Seed Weight	Maturity	Height	Lodging
LOC	3	32440501**	103.00**	1325.71**	2540.96**	61.98**
REP(LOC)	8	1514715**	3.01**	92.46**	148.31**	1.03**
FAMILY	18	448976**	25.18**	35.27**	157.45**	2.40**
LOC*FAMILY	54	172626	0.62	7.18	34.35**	0.35
FAMILY*REP(LOC)	144	145218**	0.31**	5.4**	14.84**	0.19**
GENOTYPE(FAMILY)	38	907720**	14.55**	21.57**	53.23**	4.33**
LOC*GENOTYPE(FAMILY)	114	153013**	0.40**	3.48	13.71**	0.21**
Residual	297	72867	0.19	3.24	8.52	0.09

*P-value ≤ 0.05.

**P-value ≤ 0.01.

Appendix G. Overall results of parental line and F_2 bulk performance of seed size, height, lodging, maturity, and yield of F_2 hybrids and parents in TCLP 593 2012 in replicated tests least squares means across all locations.

Family	Genotype	Yield (kg ha^{-1})	100-Seed Weight (grams 100 seeds^{-1})	Maturity (days after Oct. 1st)	Height (cm)	Lodging[§] (score)
	Female Parents					
1	NMS4-24-288	2913	8.3	24	104	3.0
2	NMS4-1-65	3480	11.4	24	87	2.3
3	NMS4-8-146	2793	9.2	25	98	2.7
4	NMS4-44-329	3086	11.0	28	95	3.5
5	NMS4-112-594	2691	7.3	22	87	3.7
6	NMS4-169-706	2906	8.9	23	86	2.6
7	NMS4-175-709	3391	11.6	24	88	2.5
8	NMS5-188-1-406	2845	11.0	21	101	2.9
9	NMS5-11-1-4	2944	10.6	24	78	2.1
10	NMS5-12-1-9	3024	11.5	24	86	3.1
11	NMS5-17-3-22	2467	11.0	25	110	3.5
12	NMS5-37-10-70	3017	11.0	26	82	2.3
13	NMS5-98-3-192	3038	11.5	26	80	2.8
14	NMS5-102-1-207	2584	8.7	27	99	3.8
15	NMS5-70-6-129	2806	9.4	26	99	3.2
16	NMS5-101-2-203	2793	10.0	27	98	3.4
17	NMS5-268-1-568	2733	6.6	24	95	3.3
18	NMS5-62-4-111	3061	9.7	27	103	3.4
19	G07-6012	3098	6.4	27	91	2.5
	Male Parents					
1	N7103	3324	8.2	25	92	2.0
2	N1703	3326	8.1	24	86	1.9
3	N7103	3404	8.3	24	91	2.0
4	N1703	3249	8.3	25	93	1.8
5	N7103	3367	8.4	24	96	1.9
6	N1703	3563	8.4	25	88	1.9
7	N7103	3406	8.3	24	82	1.8
8	N1703	3308	8.3	24	86	1.8
9	N7103	3408	8.5	24	87	1.9
10	N1703	3465	8.3	24	87	1.8
11	N7103	3364	8.3	24	90	1.8
12	N1703	3275	8.3	24	85	1.9
13	N7103	3324	8.4	25	87	1.8
14	N1703	3476	8.4	24	88	1.9
15	N7103	3386	8.3	24	91	1.9

[§]1 indicates all plants erect, 5 indicates all prostrate.

Appendix G. cont.

Family	Genotype	Yield (kg ha^{-1})	100-Seed Weight (grams 100 seeds^{-1})	Maturity (days after Oct. 1st)	Height (cm)	Lodging[§] (score)
	Male Parents					
16	N1703	3271	8.2	25	89	1.8
17	N7103	3425	8.4	25	88	1.8
18	N1703	3437	8.4	25	90	1.8
19	N7103	3594	8.4	25	90	2.0
	F_2 Bulks					
1	NMS4-24-288 x N7103	3240	8.7	25	101	2.6
2	NMS4-1-65 x N7103	3245	10.0	25	89	2.3
3	NMS4-8-146 x N7103	3046	8.6	25	98	2.1
4	NMS4-44-329 x N7103	3461	9.9	28	99	2.8
5	NMS4-112-594 x N7103	3339	7.9	24	93	2.9
6	NMS4-169-706 x N7103	3369	8.9	25	93	2.3
7	NMS4-175-709 x N7103	3460	10.2	25	93	2.2
8	NMS5-188-1-406 x N7103	3158	10.0	24	106	2.5
9	NMS5-11-1-4 x N7103	3256	9.9	26	84	1.9
10	NMS5-12-1-9 x N7103	3418	10.3	25	92	2.4
11	NMS5-17-3-22 x N7103	3097	10.1	26	120	2.6
12	NMS5-37-10-70 x N7103	3255	9.9	26	87	2.2
13	NMS5-98-3-192 x N7103	3315	9.9	27	85	2.4
14	NMS5-102-1-207 x N7103	3207	8.9	29	105	3.1
15	NMS5-70-6-129 x N7103	3269	9.3	28	103	2.8
16	NMS5-101-2-203 x N7103	3155	9.5	27	100	2.8
17	NMS5-268-1-568 x N7103	3145	7.4	25	98	2.6
18	NMS5-62-4-111 x N7103	3177	9.2	23	92	2.4
19	G07-6012 x N7103	3263	7.4	23	87	2.0

[§]1 indicates all plants erect, 5 indicates all prostrate.

Appendix H. Overall correlation matrix for yield, 100-seed weight, maturity, height, and lodging in 2012 TCLP 593 across all locations.

	Yield	100- Seed Weight	Maturity	Height	Lodging
Yield	1	0.27**	0.37**	0.00	-0.04
100-Seed Weight	.	1	0.20**	-0.70	0.15**
Maturity	.	.	1	0.43**	0.55**
Height	.	.	.	1	0.60**
Lodging	.	.	.	.	1

**P-value ≤ 0.01.

Appendix I. Plymouth, NC Location: Mean yield of parents and F_2 Hybrids in yield test TCLP 593 in 2012. Midparent (MP) heterosis (MPH), high-parent (HP) heterosis (HPH), and mean heterosis across all hybrids, in three replications.

Family	♀ Breeding Line Parent	♀ Breeding Line Parent	♂ N7103 *G. max*, HP	MP	F_2	MYE	F_2 - MYE	MPH	HPH	MPH	HPH	*G. soja* SNP Alleles[b]
		---------- Yield (kg ha^{-1}) ----------								---------- % ----------		
1	NMS4-24-288	2646	2991	2818	3004	2905	99	185	13	+7	0	31
2	NMS4-1-65	2976	2901	2939	2896	2920	-24	-43	-5	-1	0	20
3	NMS4-8-146	2843	3014	2928	2660	2971	-311	-268	-354	-9	-12	25
4	NMS4-44-329	3151	2969	3060	3145	3014	130	85	176	+3	+6	37
5	NMS4-112-594	2550	2809	2680	2986	2745	242	306	177	+11	+6	32
6	NMS4-169-706	2500	3057	2779	3211	2918	293	432*	154	+16	+5	28
7	NMS4-175-709	2967	2858	2912	3182	2885	297	269	324	+9	+11	20
8	NMS5-188-1-406	2363	3072	2718	2786	2895	-109	69	-286	+3	-9	38
9	NMS5-11-1-4	2513	3177	2845	2822	3011	-188	-22	-354	-1	-11	40
10	NMS5-12-1-9	2820	2948	2884	3022	2916	105	137	73	+5	+2	38
11	NMS5-17-3-22	2542	2723	2633	3057	2678	379	424*	334	+16	+12	34
12	NMS5-37-10-70	2475	2756	2615	3086	2685	400	471*	330	+18	+12	32
13	NMS5-98-3-192	2941	2825	2883	3305	2854	451	422	481	+15	+17	36
14	NMS5-102-1-207	2350	2914	2632	2583	2773	-190	-49	-331	-2	-11	34
15	NMS5-70-6-129	2394	2827	2611	3488	2719	769	877**	661**	+34	+23	29
16	NMS5-101-2-203	2595	2941	2768	3139	2854	285	371*	198	+13	+7	36
17	NMS5-268-1-568	2553	2905	2729	2655	2817	-162	-74	-250	-3	-9	17
18	NMS5-62-4-111	2755	2910	2832	2827	2871	-44	-5	-83	0	-3	27
19	G07-6012	2682	3182	2932	2996	3057	-61	64	-186	+2	-6	21
mean		2587	2887	2737	3015			278	128	+10[§]	+5[†]	30

*P-value ≤ 0.05.

**P-value ≤ 0.01.

[§]95% confidence interval of (2.6, 11.6).

[†] 95% confidence interval of (-2.4, 6.8).

[b]Percent alleles in female breeding line parent based on a set of 558 polymorphic SNP markers.

Appendix J. DuPont Pioneer, Kinston, NC Location: Mean yield of parents and F_2 hybrids in yield test TCLP 593 in 2012. Midparent (MP) heterosis (MPH), high-parent (HP) heterosis (HPH), and mean heterosis across all hybrids, in three replications.

Family	♀ Breeding Line Parent	♀ Breeding Line Parent	♂ N7103 *G. max*, HP	MP	F_2	MYE	F_2 - MYE	MPH	HPH	MPH	HPH	*G. soja* SNP Alleles[b]
			Yield (kg ha^{-1})							%		
1	NMS4-24-288	3506	4057	3781	3848	3919	-71	67	-209	+2	-5	31
2	NMS4-1-65	4414	4212	4313	3864	4262	-399	-449*	-348	-10	-8	20
3	NMS4-8-146	3436	3993	3714	3740	3853	-113	26	-252	+1	-6	25
4	NMS4-44-329	3634	3733	3683	4272	3708	563	588**	539*	+16	+14	37
5	NMS4-112-594	3526	3949	3737	4152	3843	309	415*	203	+11	+5	32
6	NMS4-169-706	3529	4188	3858	3589	4023	-433	-269	-598**	-7	-14	28
7	NMS4-175-709	4132	4122	4127	3876	4124	-248	-251	-245	-6	-6	20
8	NMS5-188-1-406	3626	4035	3830	3730	3933	-203	-100	-305	-3	-8	38
9	NMS5-11-1-4	3345	3874	3610	3901	3742	160	292	27	+8	+1	40
10	NMS5-12-1-9	3542	4396	3969	3865	4183	-318	-104	-531*	-3	-1	38
11	NMS5-17-3-22	2897	4097	3497	3724	3797	-73	227	-373	+6	-9	34
12	NMS5-37-10-70	3660	4167	3913	3908	4040	-132	-5	-259	0	-6	32
13	NMS5-98-3-192	3601	3929	3765	3678	3847	-168	-87	-250	-2	-6	36
14	NMS5-102-1-207	2751	3950	3350	4266	3650	616	916**	316	+27	+8	34
15	NMS5-70-6-129	3620	4011	3815	3564	3913	-348	-251	-446*	-7	-11	29
16	NMS5-101-2-203	3197	4268	3733	3787	4000	-214	54	-481*	+1	-11	36
17	NMS5-268-1-568	3417	4082	3749	4096	3916	181	347	14	+9	0	17
18	NMS5-62-4-111	3667	4088	3878	3764	3983	-219	-113	-324	-3	-8	27
19	G07-6012	3901	4275	4088	4016	4181	-166	-72	-259	-2	-6	21
mean		3547	4075	3811	3876			65	-199	+2[§]	-5[†]	30

*P-value ≤ 0.05.

**P-value ≤ 0.01.

[§]95% confidence interval of (-2.0, 6.2).

[†] 95% confidence interval of (-8.0, -1.4).

[b]Percent alleles in female breeding line parent based on a set of 558 polymorphic SNP markers

Appendix K. Athens, GA Location: Mean Yield of Parents and F_2 Hybrids in Yield Test TCLP 593 in 2012. Midparent (MP) heterosis (MPH), High-Parent (HP) Heterosis (HPH), and Mean Heterosis Across All Hybrids, in Three Replications.

Family	♀ Breeding Line Parent	♀ Breeding Line Parent	♂ N7103 *G. max*, HP	MP	F_2	MYE	F_2 - MYE	MPH	HPH	MPH	HPH	*G. soja* SNP Alleles[b]
		----------	----------	----------	Yield (kg ha^{-1})	----------	----------	----------	----------	----------	% ----------	----------
1	NMS4-24-288	3044	3350	3197	3042	3273	-232	-155	-308	-5	-9	31
2	NMS4-1-65	3580	3414	3497	3311	3455	-145	-186	-103	-5	-3	20
3	NMS4-8-146	2147	3459	2803	2668	3131	-463	-135	-791**	-5	-23	25
4	NMS4-44-329	2530	3270	2900	3394	3085	309	494*	124	+17	+4	37
5	NMS4-112-594	2547	3483	3015	3450	3249	201	435*	-32	+14	-1	32
6	NMS4-169-706	2708	3643	3175	3472	3409	63	297	-171	+9	-5	28
7	NMS4-175-709	3323	3429	3376	3545	3403	143	170	116	+5	+3	20
8	NMS5-188-1-406	2830	3186	3008	3097	3097	0	89	-89	+3	-3	38
9	NMS5-11-1-4	2861	3381	3121	3023	3251	-228	-98	-358	-3	-11	40
10	NMS5-12-1-9	2835	3163	2999	3648	3081	568	650**	485	+22	+15	38
11	NMS5-17-3-22	1821	3636	2729	2392	3182	-790	-337	-1244**	-12	-34	34
12	NMS5-37-10-70	3330	3483	3406	3224	3444	-220	-182	-258	-5	-7	32
13	NMS5-98-3-192	2673	3402	3037	3292	3220	72	255	-110	+8	-3	36
14	NMS5-102-1-207	2704	3640	3172	3119	3406	-287	-53	-521*	-2	-14	34
15	NMS5-70-6-129	3010	3690	3350	3198	3520	-322	-152	-492	-5	-13	29
16	NMS5-101-2-203	2829	3258	3044	2865	3151	-286	-179	-393	-6	-12	36
17	NMS5-268-1-568	2550	3266	2908	2713	3087	-373	-194	-552*	-7	-17	17
18	NMS5-62-4-111	2866	3543	3204	3064	3374	-310	-141	-479	-4	-14	27
19	G07-6012	2904	3645	3274	2840	3459	-620	-434	-805*	-13*	-22	21
mean		2794	3439	3117	3124			8	-315	0[§]	-9[†]	30

*P-value ≤ 0.05.

**P-value ≤ 0.01.

[§] 95% confidence interval of (-4.0, 4.7).

[†] 95% confidence interval of (-13.9, -3.9).

[b]Percent alleles in female breeding line parent based on a set of 558 polymorphic SNP markers

Appendix L. Kinston, NC Location: Mean yield of parents and F_2 hybrids in yield test TCLP 593 in 2012. Midparent (MP) heterosis (MPH), high-parent (HP) heterosis (HPH), and mean heterosis across all hybrids, in three replications.

Family	♀ Breeding Line Parent	♀ Breeding Line Parent	♂ N7103 *G. max*, HP	MP	F_2	MYE	F_2 - MYE	MPH	HPH	MPH	HPH	*G. soja* SNP Alleles[b]
			Yield (kg ha^{-1})							%		
1	NMS4-24-288	2458	2900	2679	3065	2790	276	386*	165	+14	+6	31
2	NMS4-1-65	2951	2775	2863	2938	2819	119	75	163	+3	+6	20
3	NMS4-8-146	2746	3148	2947	3114	3047	67	167	-34	+6	-1	25
4	NMS4-44-329	3027	2999	3013	3035	3006	29	22	36	+1	+1	37
5	NMS4-112-594	2139	3227	2683	2766	2955	-189	83	-461*	+3	-14	32
6	NMS4-169-706	2886	3364	3125	3204	3245	-40	79	-160	+3	-5	28
7	NMS4-175-709	3143	3217	3180	3236	3199	37	56	19	+2	+1	20
8	NMS5-188-1-406	2562	2939	2751	3018	2845	173	267	78	+10	+3	38
9	NMS5-11-1-4	3057	3199	3128	3279	3164	115	151	80	+5	+2	40
10	NMS5-12-1-9	2900	3354	3127	3136	3241	-104	9	-218	0	-6	38
11	NMS5-17-3-22	2606	3002	2804	3085	2903	182	281	83	+10	+3	34
12	NMS5-37-10-70	2605	2694	2650	2800	2672	129	151	107	+6	+4	32
13	NMS5-98-3-192	2936	3116	3026	3052	3071	-19	26	-64	+1	-2	36
14	NMS5-102-1-207	2529	3357	2943	2860	3150	-290	-83	-497*	-3	-15	34
15	NMS5-70-6-129	2199	3016	2608	2825	2812	13	217	-192	+8	-6	29
16	NMS5-101-2-203	2552	2615	2584	2829	2599	229	245	213	+9	+8	36
17	NMS5-268-1-568	2414	3445	2930	3114	3187	-73	185	-331	+6	-10	17
18	NMS5-62-4-111	2957	3208	3082	3053	3145	-92	-30	-155	-1	-5	27
19	G07-6012	2904	3247	3076	3201	3161	40	126	-45	+4	-1	21
mean		2714	3096	2905	3032			127	-64	+5[§]	-2[†]	30

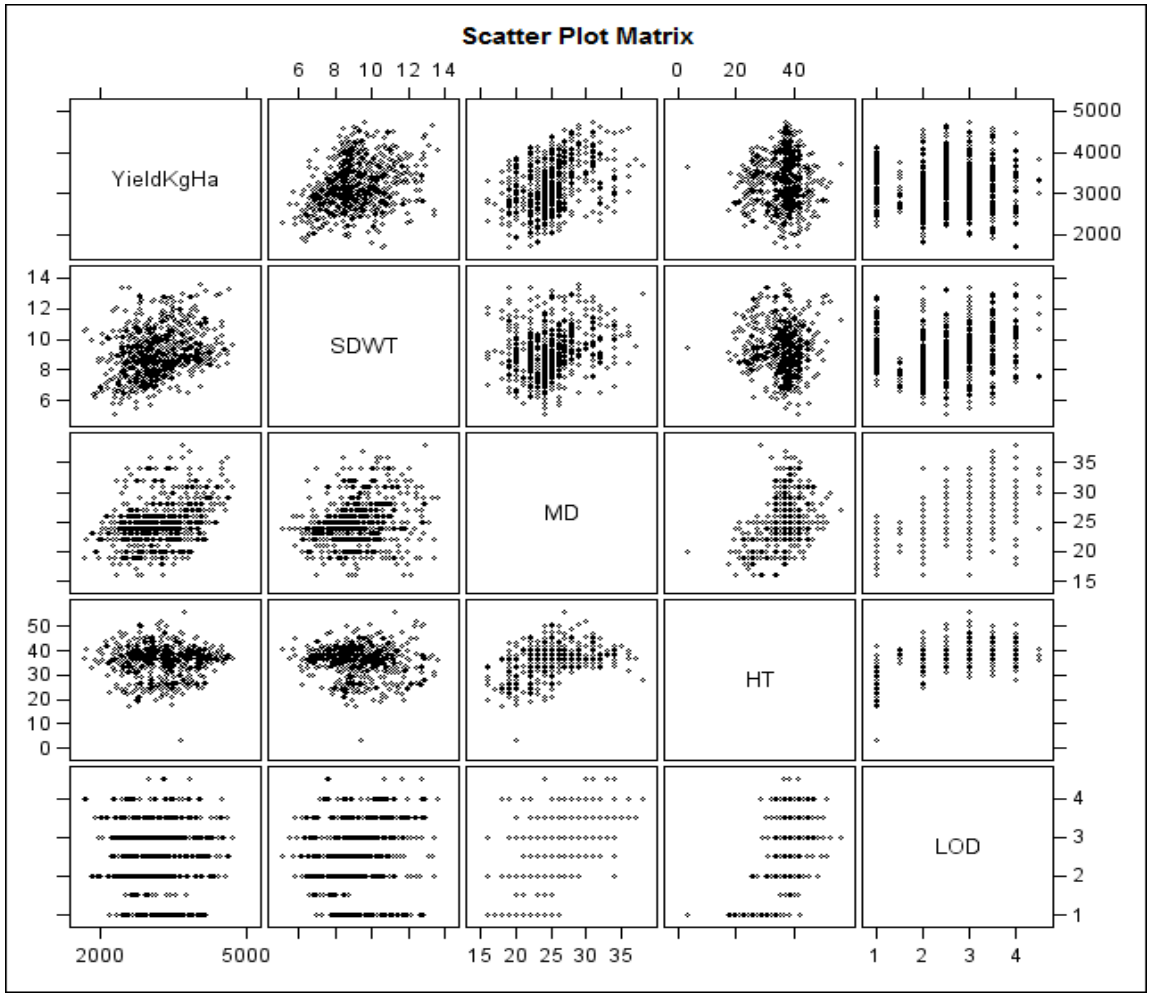
*P-value ≤ 0.05.

[§] 95% confidence interval of (2.6, 6.5).

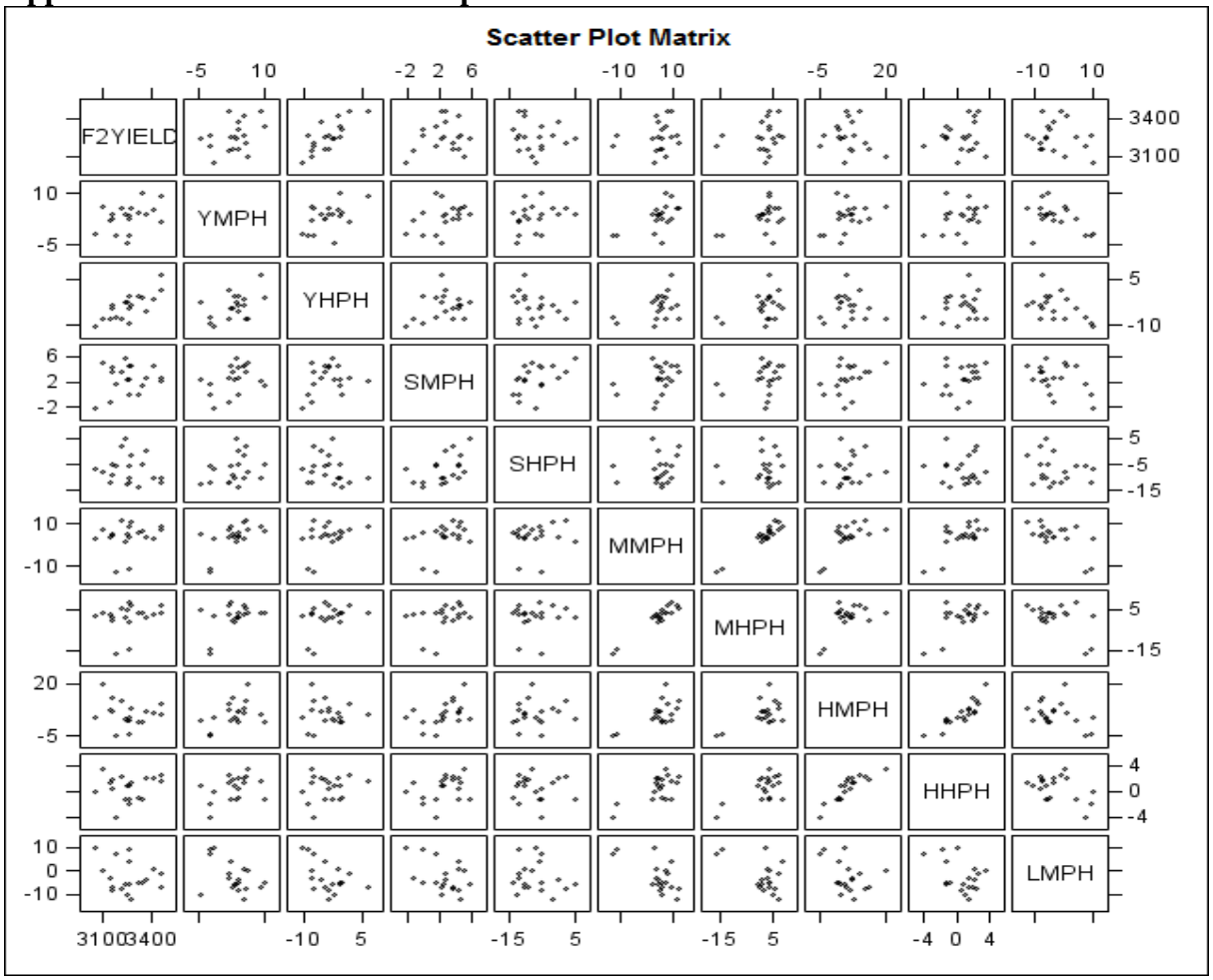
[†] 95% confidence interval of (-4.6, 1.2).

[b]Percent alleles in female breeding line parent based on a set of 558 polymorphic SNP markers

Appendix M. Correlation scatter plot matrix of the five agronomic traits measured.



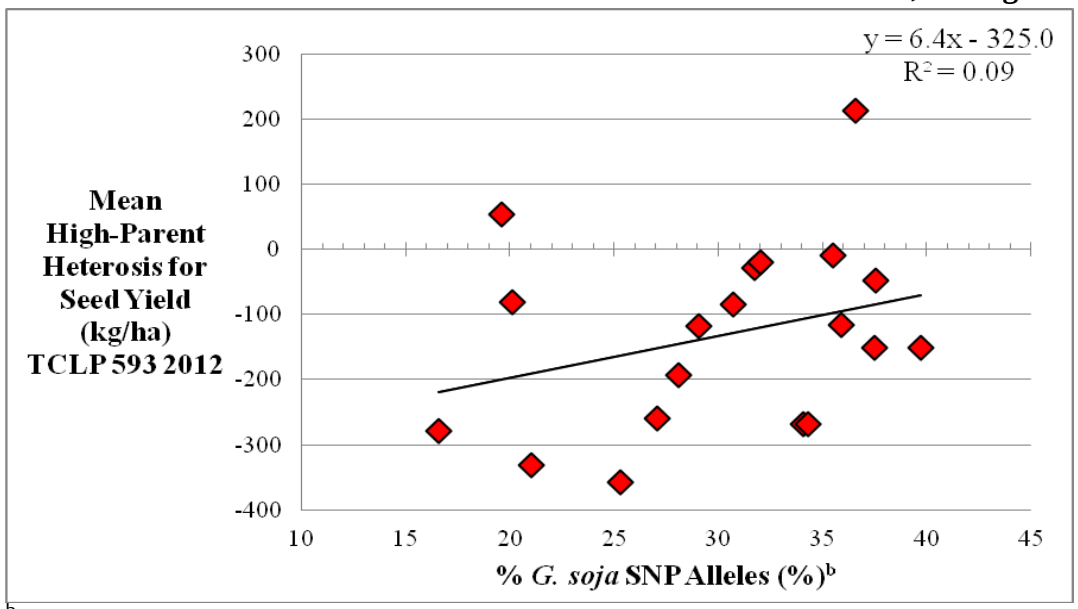
Appendix N. Correlation scatter plot matrix of ten traits measured.



Appendix O. SAS code used for analysis across all locations.

```
proc mixed data=ALL ;
class Loc Rep Fam Code;
model YieldKgHa = Fam Code(Fam)/ ddfm=kr;
random Loc Rep(Loc) Loc*Fam Loc*Code(Fam) Fam*Rep(Loc);
lsmeans Code(Fam);
```

Appendix P. Mean percent high-parent heterosis of each F_2 genotype versus percent of *G. soja* alleles in parental lines across all locations in TCLP 593 2012 with a correlation of r = 0.32; not significant.



[b]Percent *G. soja* alleles in female breeding lines based on a set of 558 polymorphic SNP markers.

Appendix Q. Mean Maturity of Parents and F_2 Hybrids. Midparent (MP) heterosis (MPH), High-Parent (HP) Heterosis (HPH), and Mean Heterosis Across All Hybrids in 2012, in Replicated Tests Averaged Across Locations.[a]

Family	♀ Breeding Line Parent	♀ Breeding Line Parent	♂ N7103 *G. max*, HP	MP	F_2	MPH	HPH	MPH(%)	HPH(%)	*G. soja* SNP Alleles (%)[b]
		----------Maturity (days after Oct. 1)----------								
1	NMS4-24-288	24.0	24.6	24.3	24.8	0.5	0.2	+2	+1	31
2	NMS4-1-65	23.7	24.1	23.9	25.2	1.3**	1.1	+6	+5	20
3	NMS4-8-146	24.8	24.2	24.5	25.2	0.7	0.4	+3	+2	25
4	NMS4-44-329	27.5	24.6	26.0	28.3	2.3**	0.8	+9	+3	37
5	NMS4-112-594	22.1	23.6	22.8	24.4	1.6**	0.8	+7	+4	32
6	NMS4-169-706	23.3	24.8	24.0	25.0	1.0*	0.2	+4	+1	28
7	NMS4-175-709	23.5	23.8	23.6	25.4	1.8**	1.7**	+8	+7	20
8	NMS5-188-1-406	21.3	24.0	22.7	23.9	1.3**	-0.1	+6	0	38
9	NMS5-11-1-4	23.8	23.7	23.8	25.8	2.1**	2.0**	+9	+8	40
10	NMS5-12-1-9	24.2	24.3	24.3	25.0	0.8	0.7	+3	+3	38
11	NMS5-17-3-22	25.3	23.6	24.5	26.2	1.7**	0.9	+7	+3	34
12	NMS5-37-10-70	26.3	23.8	25.0	26.1	1.0*	-0.3	+4	-1	32
13	NMS5-98-3-192	25.9	24.6	25.3	26.7	1.4**	0.8	+6	+3	36
14	NMS5-102-1-207	27.3	24.2	25.7	28.8	3.0**	1.5*	+12	+6	34
15	NMS5-70-6-129	25.8	24.0	24.9	27.6	2.7**	1.8**	+11	+7	29
16	NMS5-101-2-203	26.6	24.8	25.7	26.9	1.3**	0.3	+5	+1	36
17	NMS5-268-1-568	24.2	24.6	24.4	25.3	0.9	0.7	+4	+3	17
18	NMS5-62-4-111	27.1	24.7	25.9	22.6	-3.3**	-4.5**	-13	-17	27
19	G07-6012	26.5	24.5	25.5	22.7	-2.8**	-3.8**	-11	-14	21
mean		24.9	24.2	24.6	25.6	1.0**[§]	0.3**[†]	+4	+1	30

[a]Three replications in each four locations in Kinston and Plymouth, NC and Athens, GA in 2012.

*P-value ≤ 0.05.

**P-value ≤ 0.1.

[§] 95% confidence interval of (0.30, 1.73).

[†] 95% confidence interval of (-0.48, 1.03).

[b]Percent alleles in female breeding line parent based on a set of 558 polymorphic SNP markers.