OVERCOMING THE DOMESTICATION BOTTLENECK IN SOYBEAN: MAPPING DOMESTICATION TRAITS AND USING WILD SOYBEAN TO IMPROVE DIVERSITY

BY

STEPHEN ANTHONY SWARM

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Crop Sciences
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Doctoral Committee:

Professor Randall Nelson, Chair
Associate Professor Patrick Brown
Professor Brian Diers
Professor Jianxin Ma
Associate Professor Erik Sacks

ABSTRACT

QTL Mapping of Domestication-Related Traits

Soybean [*Glycine max* (L.) Merr.] was domesticated in East Asia from the wild progenitor *Glycine soja*. The domestication process led to many distinct morphological changes that adapt it to cultivation. These include larger seeds, erect growth, larger stem diameter, reduced pod shattering, and altered growth habit. The objective of this study was to identify QTL controlling key domestication-related traits (DRT). A total of 151 RILs from Williams 82 x PI 468916 and 510 RILs from Williams 82 x PI 479752 were utilized for QTL mapping. The lines were genotyped using a genotyping-by-sequencing (GBS) protocol which resulted in nearly 17,000 polymorphic SNP markers. For the eleven traits measured in this study, Haley-Knott regression was used to identify 97 QTL. The number of QTL detected for each trait ranged between 1-3 QTL in the smaller Williams 82 x PI 468916 population and 2-11 QTL in the larger Williams 82 x PI 479752 population. The majority of DRT examined in this study were controlled by minor QTL, with QTL explaining over 50% of the variation only detected for flowering date, maturity date, shattering, and pubescence type. The 97 QTL were distributed across all 20 chromosomes within 36 genomic regions. These findings identify additional QTL not detected in previous studies using smaller populations while also confirming the quantitative nature for several of the important DRT in soybeans. These results may be useful for enabling more effective use of the wild germplasm and to further understand the genetic basis of traits related to soybean domestication.

Evaluation of *G. soja*-Derived Lines

The genetic bottleneck caused during the domestication of soybean significantly reduced genetic diversity, as soybean landraces only contain half of the genetic diversity found in *G. soja* populations. Considering the narrow genetic base found within North American soybean cultivars and the greater genetic diversity present in *G. soja*, the wild species may represent a rich source of novel diversity to boost genetic gain. The objectives of this research were to identify high-yielding lines derived from *G. max* x *G. soja* crosses and characterize patterns of *G. soja* introgression in those high-yielding lines. A total of 416 *G. max* x *G. soja* lines were evaluated, all developed through the USDA soybean breeding program at Urbana, IL. A total of 26 different crosses were represented involving 14 different *G. soja* accessions and 11 *G. max* parents. In 2015, yield and other agronomic data were collected on the 416 lines at two locations

in Illinois. In 2016, 300 lines were selected and evaluated at 5 locations in Illinois, Missouri, and Nebraska. Lines which regained the yield of the soybean parent were identified within four *G. max* x *G. soja* crosses. Two lines from LN97-15076 x LG01-7770 were equivalent in yield to the *G. max* parent, LN97-15076. Twenty lines derived from Williams 82 were equivalent in yield to the soybean parent, and two lines were significantly higher yielding than Williams 82. On average, *G. soja*-derived lines contained fewer *G. soja* SNPs than what would be expected without selection. This indicates that intense phenotypic selection also led to a reduction in overall contribution from the *G. soja* parent. Within the Williams 82 x PI 479752 cross, regions containing DRT QTL tended to be in close proximity to regions with low frequencies of *G. soja* alleles. Only two lines had *G. soja* SNPs within a major shattering QTL on chromosome 16, although 20 lines contained *G. soja* introgressions within 500 kb of the QTL. High-yielding lines from this study should be good candidates to introduce novel diversity in soybean, especially lines with *G. soja* introgressions within domestication-related selective sweeps.

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Randall Nelson, for the guidance and knowledge he has shared during my studies. I would also like to thank my committee members, Dr. Patrick Brown, Dr. Brian Diers, Dr. Jianxin Ma, and Dr. Erik Sacks for their helpful advice, support, and comments. Special thanks go to our collaborators, Dr. George Graef of the University of Nebraska-Lincoln, and Dr. Andrew Scaboo of the University of Missouri, for their assistance in evaluating our *G. soja*-derived lines. These projects would not have been possible without the field and laboratory assistance of the many lab members and graduate students working with Dr. Randall Nelson, Dr. Brian Diers, Dr. Jianxin Ma, and Dr. Patrick Brown. Lastly, I would like to thank my family, especially my wife Katie Swarm, for the support and encouragement that I have received.

TABLE OF CONTENTS

CHAPTER ONE

Literature Review

Introduction and History of Soybean

Soybean (*Glycine max* (L.) Merr.) is a major world crop, ranking fourth in total harvested area behind maize, wheat and rice (FAOSTAT, 2016). Soybean has many properties that make it a useful agricultural product. As a legume, soybean has nitrogen fixing capabilities that allow it to produce yields with low nitrogen inputs. Also, the composition of the seed is approximately 40% protein and 20% oil, making it an efficient source of both edible protein and oil. The high quality and amount of protein in the seed makes soybean meal an essential component in livestock feed. This use of soybean will be become increasingly important in the future due to the rising global population. It will become increasingly important as an animal feed with diets shifting towards meat consumption and as a food in areas of the world where diets are protein poor. Production is largely dominated by five countries: United States, Brazil, Argentina, India, and China. The United States and Brazil are the two largest producers and in 2014, accounted for 33% and 30% of global soybean production respectively (USDA, 2016). In the US, soybeans were grown on 33.4 million hectares, which is only behind maize (33.6 million hectares).

Although many details of soybean domestication are debated, it has been widely accepted that cultivated soybean was domesticated from the wild annual species (*Glycine soja* Seib. & Zucc.). Most reports cite studies which conclude domestication occurred 3,000-5,000 years ago (Hymowitz, 1970; Carter et al., 2004). This estimate is mostly derived from historical data containing references to soybean in Chinese written records that appeared during the Zhou dynasty. However, other sources place domestication further back in time. Lee *et al*. (2011) studied archaeological samples from sites in China, Korea, and Japan covering a range of time periods. Some of these samples were dated as far as 5000 to 5700 years ago. Based on the size of the charred seeds, they hypothesize that domestication was already underway and place the start of domestication within 5000-9000 years ago (Lee et al., 2011). In soybean, seed weight and size appears to be controlled by multiple QTL with relatively small effects instead of a few major genes (Wang et al., 2016a; Xin et al., 2016; Kulkarni et al., 2016). It is therefore possible that soybean could have been cultivated prior to exhibiting significant changes in seed size.

Another point of debate is the location of soybean domestication. While China is generally considered to be the country of origin, multiple locations been identified as potential domestication centers based on a mixture of genetic diversity, phylogenetic, historical, and archaeological evidence. Two of the most likely origins are the Yellow River valley in north central China, and the Yangtze River valley in the south (Carter et al., 2004). Using historical records, Hymowitz proposed the Yellow River region as the center of soybean domestication (Hymowitz, 1970). Using genetic information, both regions have been identified as domestication centers. Studies using both SSRs and SNPs point to either the Yellow River valley (Wang et al., 2006; Li et al., 2008a, 2010; Han et al., 2016), or Yangtze River area (Guo et al., 2010; Wang et al., 2016b) based on levels of genetic diversity and relatedness to wild populations. While the majority of studies support a single Chinese origin, others hypothesize multiple origins within East Asia. Using archaeological samples, the argument can be made that soybeans were domesticated independently in China, Korea, and/or Japan (Lee et al., 2011; Van et al., 2014). This is largely based on the presence of large seeds in Korea and Japan at an earlier era than China. As stated previously, the quantitative genetics of seed size and the variety of uses of soybean may mean that seed size is not the optimum indicator of soybean cultivation. A study examining chloroplast DNA also came to the conclusion that soybean was domesticated in multiple centers (Xu et al., 2002). The most common chloroplast haplotype in domesticated varieties was only found in *G. soja* populations from southern China and Japan. They also found that haplotypes tended to be regionally shared between both cultivated and wild genotypes, causing the authors to conclude that soybean may have been derived from multiple domestication events.

Soybean Diversity

Although the United States is currently the largest producer of soybean, it was not grown extensively in the country until the 1940's. Soybean was first introduced to the US in 1765, by Samuel Bowen near Savannah, GA (Hymowitz and Harlan, 1983). However, production of soybean at this time was limited and short lived. The introduction of soybean to the US Corn Belt occurred at a much later time in 1851, when Dr. Edwards disseminated seeds among colleagues across several states (Hymowitz, 1987). By the early 1900's, soybean was grown predominantly as a forage crop for livestock. Grain production of soybean greatly expanded with

the onset of WWII as new edible oil sources were required to replace palm oil, and since then has expanded to become the second largest crop in the US.

While soybean has been cultivated and bred in East Asia for millennia, the North American gene pool has been largely isolated and utilizes a limited number of early introductions. This has resulted in a narrow genetic base for the commercial crop. One study investigated the pedigrees of 258 public varieties released in North America between 1947 and 1988 and identified 80 progenitor parents to the North American lines that accounted for the entire genetic base (Gizlice et al., 1994). To estimate the genetic contribution from each ancestor, the researchers calculated the coefficient of parentage based on the pedigree information. These ancestors did not contribute equally to the soybean genetic base, as only 17 ancestors accounted for greater than 1% of the genetic base and together these 17 contributed 84%. They determined that 60% of the North American genetic base can be attributed to just six ancestors—Mandarin, CNS, Richland, S-100, and the two unknown parents of Lincoln. Separate gene pools could be identified between the Northern and Southern cultivars with different ancestors involved in each region. The largest contributors to the Northern cultivars were Lincoln, Mandarin, and Richland (53% cumulative), while the Southern cultivars mainly originated from CNS and S-100 (46% cumulative). Two difficulties in using original progenitors in the pedigree analysis are that in many cases the ancestors were lost or never recorded, and potentially the pedigrees of early cultivars were incorrectly documented. To work around these issues, the authors also investigated the contributions of the first progeny from the original 80 ancestors. The narrow base of original ancestors was also reflected in the first progeny where nearly the entire constitution of Northern and Southern cultivars was determined by 70 and 46 lines respectively. This analysis illustrates the narrow genetic foundation of North American soybeans and potentially the need to introduce new germplasm into the gene pool.

The study described above had some limitations in its scope with its restriction to public releases between 1947 and 1988. A more recent study also addressed the issue of genetic diversity in North American soybeans but included a panel of cultivars more reflective of the modern genepool (Mikel et al., 2010). In this study, 2242 public and private cultivars released between 1970 and 2008 were analyzed. This is a substantially larger set of genotypes compared to Gizlice et al. (1994) and contains a different composition given only 20% of these cultivars were derived from public breeding programs. This is reflective of the shift in soybean breeding

toward private programs following the introduction of the Plant Variety Protection Act in 1970. The shift away from public breeding programs is most apparent when the most recent cultivars are considered—only 9% of releases between 2004 and 2008 originated from public programs. In one analysis, the progenitors of 494 proprietary cultivars released from 1999 through 2008 were investigated. A3127 was a major contributor to contemporary cultivars and had at least a 25% genetic contribution in 23% of the cultivars and had 10% contribution to half of the 494 cultivars. Different ancestors were prevalent in different maturity groups, with A5474, A5979 and Essex commonly used as parents in late maturity groups while MO13404 and P9061 were common in early maturity releases. Genetic diversity (0.89) calculated from the coefficient of parentage (CP) among soybean cultivars was similar to some other self-pollinating crops but was lower compared to 0.94 in rice (Wang and Lu, 2006) and 0.94 in North American dent corn (Mikel, 2008). An important consideration when using CP based on pedigrees is that the diversity is overestimated when related ancestral lines are assumed to be unrelated (Mikel et al., 2010). This study again highlights the limited scope of North American germplasm, considering the prominent use of a select number of progenitors.

While a large number of accessions were originally introduced to North America, only a small number were actually utilized as cultivars or as parents in breeding programs. Soybeans have been cultivated in China and other East Asian countries for millennia and were not subjected to a genetic bottleneck from introduction. Cui et al. (2000) conducted a coefficient of parentage analysis of 651 Chinese cultivars released from 1923 to 1995. They found that the genetic base of Chinese varieties are derived from 339 ancestors, a substantially larger number than the 80 North American ancestors. These Chinese ancestors also individually accounted for a smaller portion of the genetic base with a maximum of 7% and only 18 ancestors contributing over 1%. This is also reflected in the number of lines that contribute a cumulative of 50% of genetic base with 35 lines in the Chinese gene pool and 5 in the North American cultivars. The majority of ancestral lines originated in China, but a small number (45) were derived from foreign countries (including U.S. and Japan) that have been introduced to Chinese breeding in later decades. Unlike North America, cultivars from many different regions in China were derived from different sets of ancestors with little overlap among them. While US public breeding programs featured sharing of germplasm and cooperation in variety releases, early

Chinese breeding was largely independent and this likely contributed to the increased number of ancestors across the Chinese soybean cultivars.

The studies described above illustrate the bottleneck that occurred through the selection of plant introductions to form the North American genetic base. This can be seen by the small number of founding ancestors in North American soybeans and the larger number of ancestors contributing to Chinese cultivars. While the founder effect of introducing soybeans to North America illustrates the constrained genetic base, it was not the most influential bottleneck event in soybean's history. Three major genetic bottleneck events have been described in the history of North American soybeans—the original domestication of the crop, the selection of ancestral lines from plant introductions in the early 20th century, and intense breeding over the last 75 years. Hyten et al. (2006) studied the effects of each of these bottleneck events in a set of 25 North American cultivars, 17 North American ancestors, 52 landraces, and 26 *G. soja* accessions. By sequencing 111 fragments from 102 genes (identifying 496 SNPs), they measured genetic diversity in the populations using the proportion of pairwise differences per site, π (Tajima, 1983), and the number of polymorphic sites corrected for sample size, θ (Watterson, 1975). Of the three bottleneck events, selective breeding had the smallest effect on genetic diversity with no significant decrease in diversity detected between the elite cultivars and their ancestors. The founder effect also only led to a small decrease in genetic diversity from the set of landraces but was not found to be statistically significant. However, the North American ancestors contained only 78% of the 98 low frequency alleles (minor allele frequency $\leq$0.10) found in the Asian landraces. This is characteristic of founder effects as only a portion of the alleles are captured by the small number of founder lines with low frequency alleles least likely to persist. By far, domestication represents the largest single decrease in genetic diversity. Landraces contained a significantly lower nucleotide diversity compared to *G. soja* (66% measured by π and 49% by θ) and also a 63% lower haplotype diversity. Furthermore, domestication resulted in the largest reduction of low frequency alleles, with 81% of 237 SNPs not found in the *G. max* populations. Of the populations in the study, *G. soja* represents the largest reservoir of genetic diversity for soybean breeding.

Larger scale resequencing of *G. max* and *G. soja* genomes support the findings from the Hyten et al. (2006) study which used 102 genes. Lam et al. (2010) resequenced 31 accessions from *G. soja* (17) and *G. max* (14) at an average depth of 5x and over 90% coverage. This

represents a sizable increase in information relative to the Hyten et al. (2006) study, generating nearly 6 million SNPs in the wild accessions and over 4 million SNPs in the cultivated. The resequencing data revealed a larger amount of diversity in the wild soybean exhibited by the larger number of SNPs, higher π in the wild accessions (0.00297 compared to 0.00189), and more alleles specific to the wild population. In contradiction to the results of Hyten et al. (2006), the cultivated accessions contained a larger proportion of alleles at low frequency than the wild accessions. The authors concluded that *G. max* populations had expanded after domestication while *G. soja* populations have contracted based on an analysis of joint-allele frequency and that this would explain the differences in allele frequency. Two other observations were made relative to other major crop species. First, nucleotide diversity is lower in soybean (0.00189) compared to 0.00229 in rice (Caicedo et al., 2007), and 0.0066 in maize (Gore et al., 2009), reflecting the findings from CP analysis (Mikel et al., 2010). Second, soybean exhibits significantly higher LD due to its propensity for self-pollination. While LD decays within 30 kb in maize (Hufford et al., 2012), the average distance for LD to decay to half its maximum value in soybeans was 150 kb. However, LD decayed in half that distance in wild soybean (75 kb). These results are also consistent with another study that resequenced 302 diverse wild and domesticated soybean accessions (Zhou et al., 2015). Similar measures of genetic diversity were found for each population, with π equal to 0.00294 in *G. soja*, 0.00140 in landraces, and 0.00105 in cultivars. This again demonstrates that domestication represented the most significant genetic bottleneck with half the genetic diversity lost during the process. Using the larger sample size, LD decay estimates decreased slightly with LD extending 27 kb in wild soybean and 133 kb in soybean cultivars.

In addition to an overall decrease in genetic diversity, domestication also resulted in localized reductions in diversity at specific points of the genome. Under certain conditions, strong selection for a specific allele at a locus can lead to an extended reduction in diversity for linked loci, an event called a 'selective sweep'. Signatures of artificial selection for an allele do not always persist, with several factors affecting the degree including the level of polymorphism in the original population, the frequency of the allele prior to selection, and the amount of recombination (Walsh, 2008). Selective sweeps have been described across numerous regions in the soybean genome. By comparing six *G. soja* and ten *G. max* accessions, Chung et al. (2014) detected 206 candidate domestication regions identified by low nucleotide diversity in the

domesticated accessions relative to the wild, a sign of selective sweeps. Another study used a much larger panel of genotypes by including 105 wild and 262 cultivated accessions which were genotyped with a high-density 355,595 SNP array (Wang et al., 2016a). Based on a combination of F_{ST} and nucleotide diversity (π) calculated in 10 kb windows, they selected 1,614 windows that may have undergone selection during domestication. Although identified on all twenty chromosomes, these regions were not evenly distributed and were found in clusters 20-220 kb in size. Together, they accounted for 1.7% of the genome and 2.1% of the predicted genes in soybean. This is consistent with an earlier study which examined resequencing data across 55 accession and found that 1.5% of the genome and 2.0% of the predicted genes were under selection during domestication (Li et al., 2013). Selective sweeps have also been identified in association with specific genes. The few soybean domestication genes that have been identified to date exhibit signatures of selection in the adjacent sequences. The shattering gene *SHAT1-5* was associated with a 116 kb selective sweep in cultivated accessions caused by selection during domestication (Dong et al., 2014). Hard-seededness in soybean causes delayed imbibition of the seed and is controlled by *Gmhs1-1*. This allele arose during domestication and is also surrounded by a 160 kb sweep in *G. max* accessions (Sun et al., 2015; Zhou et al., 2015). Due to the loss of diversity and genes, selective sweeps may contain novel loci in *G. soja* that could be beneficial to soybean improvement.

Uses of Crop Wild Relatives in Breeding

The coming decades represent significant challenges to agriculture as the world population continues to rise, and both abiotic and biotic stresses evolve with changes in climate (Redden et al., 2015). To overcome these challenges, some breeders have turned to crop wild relatives (CWR) to introduce new variation. With their larger genetic diversity, wild relatives represent a valuable source of genes for crop improvement. CWR can be used for a variety of purposes, including improving abiotic stress tolerance, disease resistance, and agronomic performance. In practice, CWR are most commonly used to introgress genes for disease or pest resistance, with fewer examples of improvement in yield and quality traits (Hodgkin and Hajjar, 2008).

Wild relatives have been used in breeding across a large number of major crop species, such as tomato, wheat, and rice. Of the major crops, wheat has likely seen the most extensive use of wild relatives where they have been used to improve a wide range of traits (Maxted and Kell,

2009). Many disease resistance genes have been identified in the wild wheat species, including sources of resistance to powdery mildew (*Blumeria graminis* f. sp. *tritici*), stem rust (*Puccinia graminis* f. sp. *tritici*), leaf rust (*Puccinia triticina*), and wheat streak mosaic virus (Hoisington et al., 1999; Zhang et al., 2017). Abiotic stresses can also be improved through wild wheat species, such as drought tolerance. By introducing a chromosome segment from the wild relative *Agropyron elongatum* into cultivated wheat (*Triticum aestivum*), Placido et al. (2013) were able to improve performance under drought stress. Water stress conditions did not induce a statistically significant response in the line containing the *A. elongatum* introgression. Root and shoot biomass in the line with the *A. elongatum* translocation were significantly higher under stressed conditions compared to the wheat parental line and the negative controls (no translocation events).

Wild *Solanaceae* species also show potential in improving various traits in tomato. Late blight caused by *Phytophthora infestans* is an emerging disease in cultivated tomato that can cause heavy damage (Fry and Goodwin, 1997). Accessions from the wild species *Solanum pimpinellifolium*, *S. habrochaites*, and *S. pennellii* contain genetic resistance to this disease which can be incorporated into cultivated tomato (*S. lycopersicum*) (Ebert and Schafleitner, 2015). *S. pimpinellifolium* has been shown to be a source of qualitative resistance to late blight and has contributed three resistance genes (*Ph-1*, *Ph-2*, and *Ph-3*). All three genes have been utilized in tomato cultivars with *Ph-3* conditioning resistance to a broad range of isolates (Zhang et al., 2014). Wild tomatoes have also been used for improving quality traits. In tomatoes, high soluble solid content (high sugar concentration) is a valuable trait at the processing stage. Two examples of using CWR to increase sugar content in tomato originated from crosses with *S. chmielewskii* (Rick, 1974) and *S. penellii* (Fridman et al., 2004). Using the green-fruited *S. chmielewskii*, soluble solids were raised two percentage points above the *S. lycopersicum* recurrent parent. However, the gain in sugar concentration was also correlated with a few deleterious traits such as fruit cracking and softness (Rick, 1974). To fine map QTL from another species, *S. penellii*, Fridman et al. (2004) used an advanced backcross population. One QTL, *Brix9-2-5* was responsible for an 11-25% increase in Brix and mapped to a cell-wall invertase gene.

In rice (*Oryza sativa*), the wild species *O. rufipogon* has been used to improve yield and agronomic traits. To map QTL for yield-related traits, a set of 159 introgression lines was created

containing segments from *O. rufipogon* (Tian et al., 2006). Although the cultivated species is superior to *O. rufipogon* in yield and associated traits, they were able to find six QTL in which the *O. rufipogon* conferred an improvement in three traits (panicles per plant, grains per panicle, filed grains per plant). Progeny derived from wild rice have also been released as improved cultivars (Ram et al., 2007). The variety Dhanrasi was derived from crosses with *indica* rice cultivars and an *O. rufipogon* accession, which contributed 25% of the pedigree. Dhanrasi was an improvement over the *O. sativa* parents in both blast resistance and yield, which indicates yield-enhancing genes may have been contributed by *O. rufipogon* in addition to its disease resistance.

Uses of CWR can also be found in soybean. Within the *Glycine* genus, two subgenera have been defined as *Glycine*, which contains the perennial species, and *Soja* composed of the annual species. Within the *Glycine* subgenus, there are currently 26 species all native to Australia (Sherman-Broyles et al., 2014). These perennial species fall within the tertiary genepool classification given by Harlan and de Wet (1971) due to the natural crossing barriers that prevent hybridizations with *G. max* and require extreme measures to produce crosses. Very limited success has been made in hybridizing species from the *Glycine* subgenus with cultivated soybean. Of the 26 species, *Glycine tomentella* has been shown to be the most promising for hybridizations with soybean (Ladizinsky et al., 1979). Successful crosses with made between the *G. tomentella* accession PI 441001 and the soybean cultivar 'Dwight' (Singh and Nelson, 2015). *G. tomentella* and other *Glycine* perennials may benefit soybean breeding in the future due to their resistance to various diseases and tolerance to abiotic stresses (Sherman-Broyles et al., 2014).

Soybean germplasm is maintained in several collections worldwide, and although the *Glycine* subgenus is greater in terms of species, the *Soja* subgenus is far larger in the number of accessions (Carter et al., 2004). The subgenus *Soja* is comprised of the two annual species, *G. max* and *G. soja*. These two species form a primary genepool, and therefore the research using CWR in soybeans has almost exclusively utilized *G. soja*. Due to the diversity lost during the domestication bottleneck and selective sweeps, *G. soja* contains novel alleles and genes that are not present in the current soybean genepool. This is exemplified by both genetic diversity studies and a pan-genome analysis. Across multiple studies, *G. soja* populations have been found to contain higher genetic diversity than cultivated populations as measured by allele richness of

SSR markers (Guo et al., 2010) and nucleotide diversity of SNPs (Hyten et al., 2006; Lam et al., 2010; Zhou et al., 2015). In addition to genetic markers, differences between wild and domesticated populations can also be seen when examining genes and whole-genome sequencing. Zhou et al. (2015) resequenced 302 wild and cultivated accessions and aligned the sequences to both the *G. max* reference genome (Williams 82) and a *G. soja* genome (IT182932). By aligning to the *G. soja* genome, additional annotated resistance genes were identified in the wild accessions, of which only half of these were also present in the cultivated accessions (Zhou et al., 2015). By comparing seven *G. soja* genomes to the Williams 82 reference genome (Glyma.Wm82.a1.v1.1), Li et al. (2014) identified sequences and genes specific to the wild genomes. On average, the *G. soja* accessions contained nearly 1400 more genes compared to the Williams 82 genome and 2.3-3.9 Mb of presence-absence variants specific to *G. soja*.

These differences between *G. max* and *G. soja* have on occasion made the wild species a useful source of new variation. *G. soja* is usually used to improve those traits that can be evaluated per se in the wild species—traits such as disease resistance, abiotic stress tolerance, and quality traits. One disease that can be combated using *G. soja* is *Heterodera glycines*, commonly known as soybean cyst nematode (SCN). This nematode is one of the most damaging soybean pathogens in the United States, and currently the most efficient management practice is to use genetic resistance in conjunction with crop rotation (Niblack et al., 2006). Only a few sources of resistance have historically been used in soybean varieties, and these genes are vulnerable to losing their efficacy through evolving pathogen populations (Kim et al., 2011). Using the *G. soja* accession PI 468916, Wang et al. (2001) identified two *G. soja* QTL that contributed to resistance to SCN HG type 2.5.7. Two strong QTL were located on chromosome 15 (*cqSCN-006* explained 23% of the variation) and chromosome 18 (*cqSCN-007* explained 27%). Although *rhg1* is also located on the same chromosome as *cqSCN-007*, the two QTL are separated by 61 cM and therefore were determined to be unique (Kim and Diers, 2013). When combined with other resistance sources (*rhg1-b* from PI 88788 with *rhg1* and *Rhg4* from PI 437654), these *G. soja* QTL significantly increased resistance to three SCN isolates (Kim et al., 2011). This indicates that the *G. soja* source of resistance would be a valuable tool to bolster resistance in soybean cultivars and a germplasm line containing the PI 468916 SCN resistance has already been released (Diers et al., 2005).

Given the broad range in distribution, *G. soja* has adapted to a wide variety of adverse growing conditions. Several examples of abiotic stress QTL in *G. soja* have been described in the literature, including salt tolerance (Zhu et al., 2012; Qi et al., 2014), drought tolerance (Luo et al., 2013; Tang et al., 2013), and cold tolerance (Yang et al., 2014). A novel candidate gene for salt tolerance was recently identified in *G. soja* using whole-genome sequencing information along with an interspecific RIL population (Qi et al., 2014). Using this population, they were able to map the salt tolerance QTL to a 388 kb region on chromosome 3. One gene, *Glyma03g32900*, was highly similar to cation H^+ exchangers and contained a retrotransposon insertion that was absent in the homologous *Glysoja01g005509* gene (*GmCHX1)* in the *G. soja* genome. This *GmCHX1* gene could potentially be utilized in breeding to improve salt tolerance of soybean cultivars. To identify additional regions that likely contain abiotic stress QTL, Anderson et al. (2016) performed an environmental association analysis in *G. soja* accessions. This analysis utilizes the geographical data from accessions in the USDA germplasm collection to detect adaptations in wild populations for specific environments. Some SNPs associated with environmental variables were unequally represented in wild and domesticated populations. For example, the most significant SNP associated with average wet season temperature was found in 90% of *G. soja* accessions but only 3% of soybean cultivars. This could be indicative of genes unique to wild accessions that can be used to improve abiotic stress tolerance in soybean.

Wild soybean also carries QTL that may be used to increase protein concentration in cultivated varieties. Seeds from *G. soja* accessions commonly contain a larger proportion of protein compared to *G. max*. In one survey, *G. max* accessions produced 366-429 g kg^{-1} compared to 418-506 g kg^{-1} in *G. soja* (Chen and Nelson, 2004). Using the PI 468916 accession, Diers et al. (1992) identified two QTL that increased protein yield by 24 g kg^{-1} and 17 g kg^{-1}. After backcrossing the QTL into the A81-356022 *G. max* background and further testing, it was determined that only the QTL on LG I (chromosome 20) retained a significant effect at 21 g kg^{-1} (Sebolt et al., 2000). This QTL was also introduced into three other *G. max* backgrounds, and the effect of the *G. soja* allele was not significant in the high protein C1914 background. The authors concluded that the allele was therefore not unique to PI 468916 and could also be found in some *G. max* germplasm (Sebolt et al., 2000). Fine mapping of this locus delimits the location to a 3-cM interval and was also associated with an inverse relationship with other agronomic traits, including oil concentration and yield (Nichols et al., 2006).

Glycine soja and other CWR are most commonly used to improve traits that can be identified in the wild accession, but they can also be utilized to improve traits that are unobservable per se, such as seed yield. When using *G. soja* parents, multiple backcrosses are required to efficiently regain the soybean phenotype for important agronomic traits due to the abundance of undesirable traits in the wild soybean (Carpenter and Fehr, 1986). One recent study compared *G. soja*-derived lines to their *G. max* parent in order to identify improved yields (Akpertey et al., 2014). Four wild accessions from China and one accession from Russia were crossed to Williams 82 to form BC_1 lines. These lines were then either backcrossed to Williams 82 to form BC_2, crossed with another BC_1, or crossed with a more modern cultivar (IA2052, IA3023, or LN97-15076). While none of the *G. soja*-derived lines exceeded the higher yielding recurrent parent, some lines were not significantly different from the *G. max* parent.

The inability to improve upon the yield of the soybean parent is likely due to the complex nature of the trait and is not indicative of an absence of beneficial alleles in *G. soja*. In fact, QTL from wild soybean have been mapped which improve yield. Concibido et al. (2003) reported the identification of a yield improvement QTL originating from PI 407305. This QTL was mapped to a region on LG B2 (chromosome 14) using an advanced backcross method with 265 BC_2 individuals. In the mapping population, the individuals carrying the PI 407305 allele were significantly higher yielding in 6 out of 10 environments. The QTL was then validated in a variety of elite genetic backgrounds from Asgrow germplasm. A significant increase was only observed in two out of six backgrounds, where it was responsible for a 5-9% increase in yield. However, the remaining backgrounds were only tested in one environment and more extensive testing may reveal similar effects in yield. In spite of the genetic background limitations, this study demonstrates the potential for using *G. soja* to increase yield. Another study successfully mapped a seed yield QTL using the *G. soja* accession PI 245331 backcrossed to the *G. max* variety '7499' (Li et al., 2008b). Mapping in a population of 147 BC_2F_4 lines identified a QTL for yield on LG A1 (chromosome 5) which accounted for 12% of the variation and had an additive effect of 223 kg ha^{-1}. A second population derived from the same BC_2F_1 plants was also used to validate the QTL effects. BC_2F_4 progeny in the validation population that were homozygous for the *G. soja* allele at the QTL yielded 6.3% higher than lines homozygous for the 7499 allele. This is another example of the potential utility of using *G. soja* to improve yield in soybean.

Domestication Related Traits

An obstacle to utilizing CWR for crop improvement is the persistence of phenotypes that hinder cultivation. The process of domestication from a crop wild relative leads to several morphological and physiological transitions that are consistent across most species. This set of traits that are associated with domestication are commonly referred to as the 'domestication syndrome' (Hammer, 1984). Compared to their wild ancestors, domesticated food crops are typically characterized by larger fruits or grains, loss of seed dispersal and seed dormancy, increased apical dominance, and more robust plant growth (Doebley et al., 2006). Domesticated plants also exhibit more synchronized flowering between male and female flowers/organs and more even grain ripening (Harlan, 1995). In many cases, the extent of this domestication syndrome can be so severe that the species requires human cultivation to persist, such as maize which entirely lacks mechanisms for seed dispersal. Other crops, such as carrots, are able to persist without human involvement (Doebley et al., 2006). Genetic dissection of these key domestication traits can be an important tool to better understand the mechanisms controlling the important traits for cultivation. Breeders can then use this information to efficiently recover the domesticated phenotypes using marker assisted selection. Alternatively, regions adjacent to loci related to domestication can be targeted for introgression from the wild species. This approach may be useful for identifying beneficial genes lost during domestication through selective sweeps (the reduction of diversity in regions linked to a causative mutation due to selection).

Studies have been conducted to understand domestication-related traits in many species, including maize, rice, and tomato. In maize, genes have been identified that control two traits essential to domestication—apical dominance and the loss of the protective kernel casing. The effects of apical dominance are easily seen when comparing domesticated maize (*Zea mays* ssp. *mays*) with its wild ancestor, teosinte (*Zea mays* ssp. *parviglumis*). Teosinte exhibits long lateral branches originating from nodes along the main stem with terminal male inflorescences (tassels) on each branch. In contrast, domesticated maize features a single main stem with a terminal tassel and only two or three short lateral branches tipped by female inflorescences (Doebley et al., 1995). One locus, *teosinte branched1* (*tb1*),was found to contribute a large effect on branching architecture and was mapped to the long arm of chromosome 1 (Doebley and Stec, 1993; Doebley et al., 1995). Plants with the *tb1-ref* allele produced a number of basal tillers and lateral branches similar to teosinte. Doebley identified the underlying gene at this locus using the

Mutator transposable element system and determined that the encoded protein is homologous to *cycloidea* in snapdragon (Doebley et al., 1997). In addition to their sequence homology, the genes also share some functional features. The snapdragon *cycloidea* represses axillary growth (flowers) and is involved in floral organ identity by hindering stamen growth. Likewise, maize *tb1* represses axillary growth (branches) and determines the sex of the terminal inflorescences on lateral branches. In addition to the development of apical dominance, the loss of encased kernels was a key feature of maize domestication. Teosinte exhibits small seeds that are completely enclosed in a hard, protective cupule composed of the rachis internode and outer glume. Maize kernels are only partially contained in shallow cupules in which the outer glume simply cups the exposed kernel instead of enveloping the seed. A major locus, *teosinte glume architecture 1* (*tga1*), was mapped to a region on chromosome 4 and accounted for 42-50% of the variance in glume structure (Dorweiler et al., 1993). Fine-mapping later determined that the locus lies in a 6-kb region with a gene homologous to SBP transcription regulators (Wang et al., 2005). The change in phenotype between the teosinte and maize alleles is due to a SNP leading to a non-conservative amino acid substitution, likely affecting protein function. Maize plants with *tga1* alleles mutated with EMS at the SBP gene recovered the teosinte phenotype with enlarged glumes encasing the kernel, thereby confirming the role of the SBP gene as *tga1*.

As one of the most widely cultivated crops globally, domestication of rice has been extensively studied including the dissection of traits such as shattering and seed dormancy. In wild rice species (*Oryza rufipogon* and *O. nivara*), ripe seeds are easily dispersed through shattering whereas cultivated rice (*O. sativa*) is much more resistant. Two major genes have been identified that significantly reduce the susceptibility to shattering, *sh4* and *qSH1*. The *sh4* locus had been identified on chromosome 4 through QTL mapping in populations developed from crosses with cultivated and wild rice (both *O. rufipogon* and *O. nivara*). Several QTL were detected in these populations, but the *sh4* locus consistently explained the majority or a large portion of the variance and the wild allele displayed dominant gene action (Xiong et al., 1999; Li et al., 2006a). The *sh4* gene was found to be a Myb3 transcription factor that is involved in the development of the abscission layer between the grain and pedicel (Li et al., 2006b). The shattering prone *O. nivara* accessions displayed a fully developed abscission layer of thin-walled cells. Shattering resistance in *O. sativa* is associated with discontinuity in this cell layer by the replacement with thick-walled cells. Variation for grain shattering also occurs within *O. sativa*

and appears to follow a correlation with the degree of discontinuity in the abscission layer. Cultivars in the *japonica* group tend to be more resistant to shattering than *indica* cultivars. When Li et al. (2006b) compared the abscission layers between two *O. sativa* cultivars, they found the *indica* cultivar with more shattering had a more complete abscission layer and developed closer to the vascular bundle. Another major shattering gene, *qSH1*, was identified by utilizing this variation in shattering within domesticated rice, *O. sativa*. Using a population from the cross of a shattering susceptible *indica* cultivar with a non-shattering *japonica* cultivar, *qSH1* was mapped to a region on chromosome 1 and explained 68% of the variation in the population (Konishi et al., 2006). A near-isogenic line (NIL) containing the shattering *qSH1* allele had a strong shattering phenotype and a complete abscission layer (compared to the shattering resistant genotype, which lacked an observable abscission layer). Fine mapping identified an ortholog to *Arabidopsis REPLUMLESS* (*RPL*), which is involved in abscission layer formation in *Arabidopsis*. With changes at either loci (*qSH1* or *sh4*), the severity of shattering decreases but does not completely prevent seed dispersal. Selection for shattering resistance likely occurred due to the persistence of resistant genotypes in harvested seeds and loss of shattered genotypes. Complete resistance to shattering can be a detriment to cultivation as harvesting techniques rely on the separation of the grain from the pedicle with threshing, which maintains the presence of shattering QTL within domesticated *O. sativa* (such as *qSH1*).

Seed dormancy is important in the wild species as it delays germination of the seed until sustained conditions for growth occur. In agriculture however, reliable and uniform germination is important after sowing of seeds. Rice is no exception to this rule where domesticated rice exhibits significantly higher germination rates compared to the wild species. Using interspecific crosses, some QTL have been mapped which control the loss of dormancy from the wild ancestor, although to date no genes have been cloned from these QTL. A number of QTL were detected in crosses with perennial *O. rufipogon* (Cai and Morishima, 2000) and annual *O. nivara* (Li et al., 2006a), however only a QTL on chromosome 6 was consistent between the two studies. In spite of its importance in domestication, rice cultivars also exhibit varying degrees of seed dormancy. Like shattering, complete loss of seed dormancy can be harmful for agriculture due to pre-harvest sprouting while the grain is still attached to the panicle. To investigate the genetics controlling dormancy, QTL mapping was performed in a population derived from crosses with a *japonica* cultivar with low seed dormancy and an *indica* cultivar with desirable

levels of dormancy to resist pre-harvest sprouting (Lin et al., 1998). Five QTL were detected in this study, with the strongest effect derived from a locus on chromosome 3 which explained 25% of the variation. Molecular cloning of this QTL (*Sdr4*) revealed a gene for a protein with unknown function but may operate as an intermediate regulator of dormancy during seed maturation (Sugimoto et al., 2010). Haplotype analysis was conducted across both cultivated and wild rice accessions to investigate the role of *Sdr4* in domestication. They found three haplotypes (*Sdr4-n, Sdr4-k, Sdr4-k'*) across 59 *O. sativa* cultivars, which only occurred in three *O. rufipogon* accessions (*Sdr4-k* and *Sdr4-k'* which were associated with decreased germination rate in *O. sativa*). The divergence in this locus between the domesticated and wild species may indicate that this region was under selection during the domestication process.

Among the traits associated with the domestication syndrome, fruit size represents one of the few traits that may have been subject to intentional selection by humans (Ross-Ibarra et al., 2007). This trait is also under continued selection by modern breeding which seeks to improve crop yields. The dramatic increase in fruit weight following domestication is especially apparent in tomato. The wild progenitor (*Lycopersicon esculentum*) has small fruits weighing a few grams, whereas cultivated tomatoes may weigh 1000 grams (Frary et al., 2000). As a quantitatively inherited trait, many QTL control fruit weight in tomato but one QTL, *fw2.2*, is associated with up to a 30% increase in weight. Wild species of *Lycopersicon* predominantly carry the small-fruit allele of *fw2.2* while cultivated tomatoes carry the large-fruit allele, indicating that this locus may have been important during the domestication process (Alpert et al., 1995). The underlying gene for *fw2.2* was identified as *ORFX*, but homologous genes in other species had unknown functions. Analysis of the protein structure indicates that ORFX shares features of the RAX family, which includes proteins controlling cell division similar to how *ORFX* controls carpel cell number (Frary et al., 2000). The effect on fruit size appears to be controlled not by a change in sequence of the gene, but instead by a change in gene regulation. Specifically, the *fw2.2* alleles differ in their timing of expression—peak expression of the large-fruit allele occurred approximately one week earlier than the small-fruit allele (Cong et al., 2002). The large-fruit alleles is also associated with a more gradual and sustained increase in cell division, compared to the small-fruit allele which leads to a rapid but brief increase in mitosis. The *fw2.2* gene was one of the first examples of a heterochronic change regulating evolution in plants. Unlike some other domestication genes (e.g. *tb1*), *fw2.2* confers an incremental change

instead of a dramatic shift in phenotype. The increase in fruit size during domestication was likely due to the stacking of *fw2.2* along with many other loci (Clint Nesbitt and Tanksley, 2002).

Mapping Domestication Traits in Soybean

Although *G. max* and *G. soja* are genetically similar, their morphology is quite diverged. Cultivated soybean is characterized by upright growth, short stature (including determinate growth in many varieties), and large yellow seeds. Wild soybean, on the other hand, exhibits creeping or climbing growth, small black seeds (approximately 1.5 g/100 seeds), dehiscent pods, and prolific branching. These traits are all key to the domestication of soybean and are required for efficient cultivation. While soybean is a globally important field crop, relatively little research has been conducted to identify QTL or genes associated with domestication related traits in soybean. An early study that mapped traits in an interspecific population was conducted to study variation in several quantitative traits (Keim et al., 1990a). A biparental recombinant inbred line (RIL) population of 62 genotypes was created from a *G. max* (A81-356022) by *G. soja* (PI 468916) cross. The F_2-derived F_3 progeny were grown in 1.5 m rows and evaluated for eight quantitative traits (R1, R8, seed-fill length, stem diameter, stem length, canopy height, leaf width, and leaf length). Using 150 RFLP markers, they detected significant marker-trait associations for all traits (between 1 and 5 associations detected for each trait), with the markers explaining 16-24% of the total phenotypic variation. Some markers were associated with multiple traits, including within and between reproductive and vegetative traits.

Another study was conducted in Japan to investigate the genetic basis controlling domestication traits (Liu et al., 2007). To do this, they made a RIL population between the *G. soja* line Hidaka 4 and the *G. max* line Tokei 780. Phenotyping was conducted in the greenhouse for two seasons, and several traits distinguishing the two parents were measured—flowering time, determinate habit, plant height, number of nodes, maximum internode length, twinning habit, pod dehiscence, seed weight, and hard seededness. Distributions of the traits showed a normal distribution between the two parents for most of the traits, except flowering time and determinate habit (bimodal distributions), and hard seededness (two modes, indicating a single major gene). QTL mapping was conducted on the 96 RILs with 282 SSR and isozyme markers using composite interval mapping. Fourteen significant QTL were detected for the domestication related traits and were found across six linkage groups. Most traits were found to be controlled

by only one or two major QTL, and some QTL were significant for multiple traits. A single major QTL was detected for both determinate habit (LG L (chromosome (chr) 19), near the location of *Dt1*) and pod dehiscence (LG J, chr 16). Flowering time appeared to be controlled largely by a QTL on LG C2 (chr 6), which correlates to the *E1* maturity gene. This region also showed an association with hard seededness, in addition to a highly significant region on LG D1b (chr 2). Growth traits were associated with regions on LG C2 (chr 6) (maximum internode length), D1b (chr 2) (twining habit), G (plant height, number of nodes, maximum internode length, and twining habit), and L (maximum internode length). The more numerous QTLs for these traits are a reflection of their quantitative nature, and the common region on LG G (chr 18) may be indicative of the correlations among some of the traits.

Another study mapping traits in a *G. max* x *G. soja* population was conducted by Wang *et al.* in 2013 (Wang et al., 2013). The main objective of this study was to develop a chromosome segment substitution line (CSSL) population using *G. soja*, while also testing the potential usefulness of the population by mapping two traits related to domestication (plant height, and the number of nodes on main stem). Their CSSL population was composed of 151 backcross lines and recovered 95.7% of the wild soybean genome. They identified four chromosome segments from the wild parent that were associated with plant height, found on chromosomes 4, 10, 12, and 19. The QTL/segment found on 19 appears to be consistent with a QTL also found by the Liu et al. (2007) study and correlates to the gene for determinate growth, *Dt1*. This was also the largest effect QTL in the Wang et al. study, accounting for 42% of the phenotypic variation. Four segments were also significantly associated with the number of nodes on main stem, located on the same chromosomes as plant height (chromosomes 4, 10, 12, 19). The segments on chromosomes 10 and 19 conferred pleiotropic effects while the segments on 4 and 12 were located on separate regions from the corresponding QTL for plant height.

To date, only a few domestication-related traits have been narrowed from a QTL region to identifying the underlying gene, including determinate growth, pod shattering, and hard-seededness. Indeterminate growth is characterized by the continuation of apical stem growth after initiation of flowering. While both indeterminate and determinate growth types can be observed in cultivated soybean, wild soybeans are exclusively indeterminate. Based on classical genetics, this trait has been known to be controlled by the single gene *Dt1* (Woodworth, 1932) on chr 19 (LG L). Map-based cloning and comparative genetics led to the identification of the

underlying *Dt1* gene, an ortholog of *Tfl1* which controls determinacy in *Arabidopsis* (Liu et al., 2010; Tian et al., 2010). The recessive determinant phenotype was found to be conferred by one of four possible nonsynonymous SNP mutations. These four *Gmtfl1* alleles were observed in *G. max* genotypes but were entirely absent from the *G. soja* accessions that were screened, which supports the absence of an observable determinant phenotype in wild soybean (Tian et al., 2010). The presence of determinate alleles in a diverse set of Chinese landraces suggests that selection for the determinate phenotype occurred prior to the local adaptation of these genotypes (Tian et al., 2010). Whether this trait arose during domestication or shortly after is unknown.

Pod shattering has long been studied in soybean due to its effect on yield loss. While elite cultivars are typically largely resistant to shattering, landrace varieties exhibit varying degrees of shattering with wild soybean accessions displaying the highest susceptibility. Several regions have demonstrated associations with shattering tendency, but a QTL on LG J (chr 16) (*qPdh1*) was most consistently detected and explained the largest amount of variation (Bailey et al., 1997; Liu et al., 2007; Suzuki et al., 2009). This QTL was fine mapped to a 134-kb region on chromosome 16 containing 10 predicted genes. None of the identified putative genes are orthologous to *Arabidopsis* genes associated with pod dehiscence, however a dirigent-like protein (*Glyma16g25580*) in this region has been identified as *Pdh1* (Funatsuki et al., 2014). Shattering-resistant plants, when transformed with the susceptible allele of the dirigent-like protein, displayed significantly increased shattering and torsion. The group found that the *Pdh1* transcript was abundant in the inner sclerenchyma of pod walls and likely promotes dehiscence by increasing torsion forces through lignin accumulation. Another gene related to pod dehiscence was cloned to a separate region of chr 6 and named *SHAT1-5* (Dong et al., 2014). The researchers in this studied noticed thickening in the *G. max* fiber cap cells (FCC) relative to *G. soja*, indicating that cell wall thickening in the FCC is related to shattering resistance. Soybean genes orthologous to *Arabidopsis* genes regulating lignified valve margin cells were cloned, including the NAC gene *Glyma16g02200*. This gene was determined to be *SHAT1-5* due to its location within a shattering QTL, signatures of selection in domesticated soybean, and RNA expression patterns. Shattering resistant genotypes are characterized by heavily lignified FCC in the ventral suture, accompanied by increased expression (15-fold) of *SHAT1-5* in the FCC. Together *SHAT1-5* and *PDH1* may interact to prevent pod dehiscence in cultivated soybean, but further work will be required to fully explain the processes involved.

Another domestication-related trait that has been located to a single gene is hard-seededness. Genotypes with the hard-seeded phenotype exhibit delayed imbibition, a trait beneficial for the survival of wild species but detrimental to cultivation. A QTL controlling hard-seededness has been mapped to a region of chromosome 2 across several mapping studies, including in *G. max* x *G. soja* populations (Keim et al., 1990b; Liu et al., 2007). Fine mapping in a study using *G. max* x *G. soja* plants narrowed the region to a 22-kb segment with only two genes (Sun et al., 2015). Of the two genes, *Glyma02g43700* contained a single mutation between the *G. max* and *G. soja* parents. Expression of the gene (*GmHs1-1*) was primarily localized in developing seed coats, and transformations of a permeable soybean cultivar with the *G. soja* gene sequence resulted in impermeable seeds. While the cellular mechanism controlling seed coat impermeability is not understood, *Gmhs1-1* is believed to encode a calcineurin-like metallophosphoesterase protein and the accumulation of calcium in the seed coat is associated with impermeability. When accessions were screened for the *GmHs1-1* locus, all hard-seeded wild accessions carried the *GmHs1-1* allele and soybean cultivars with permeable seeds carried the *Gmhs1-1* allele. Interestingly, some landraces carried the impermeable allele while avoiding the hard-seeded phenotype through high frequencies of seed coat cracking. The selection for permeable seed coats during domestication led to a 160-kb selective sweep around *GmHs1-1*.

The studies discussed thus far have followed a similar 'top-down' approach to finding loci related to domestication in which QTL are mapped based on a phenotype and potentially leads to the identification of candidate genes (Ross-Ibarra et al., 2007). Population genetics can then be used to confirm that the gene was tied to the domestication process. An alternative 'bottom-up' approach begins by identifying regions with signatures of selection in a population from which loci can then be linked to phenotypes. While 'top-down' approaches have been traditionally used to identify genes or QTL, population genetics are increasingly being used to identify loci of interest, including in soybean (Li et al., 2013; Chung et al., 2014; Zhou et al., 2015, 2016; Han et al., 2016). These studies compare genomic data from panels of wild, landrace, and improved soybean accessions to identify regions that underwent selection during domestication. Various statistics can be used to detect these selection signals including cross-population composite likelihood ratio (Zhou et al., 2015), F_{ST} (Han et al., 2016), Tajima's D (Li et al., 2013), and reduction of diversity (ROD) (Chung et al., 2014; Zhou et al., 2016). While these statistics differ in the exact properties that are measured, they all are used to detect regions

that exhibit significant differentiation between populations in a departure from the neutral theory model. Regions of low polymorphism in the domesticated accessions relative to the wild could be indicative of selective sweeps that contain genes under selection during domestication. All of these studies were able to identify between 140 to 382 candidate domestication regions across all 20 chromosomes. Using 302 accessions, covering 110 improved cultivars, 130 landraces, and 62 *G. soja* accessions, Zhou et al. detected 121 selective sweeps and 81 copy number variants selected during domestication (Zhou et al., 2015). The authors compared these selective sweeps with the results from a QTL mapping study of domestication-related traits (Liu et al., 2007) and found that nearly all of the detected QTL overlapped with selective sweeps. The 'bottom-up' approach in this study presents an advantage over the traditional QTL mapping in that the regions defined by the selective sweeps were much narrower than the confidence intervals for the corresponding QTL. For example, the pod dehiscence QTL from Liu et al. (2007) covered a confidence interval of 8.7 Mb on chromosome 16, while the selective sweep only covered a 190 kb range with 14 genes. The *Pdh1* gene reported by Funatsuki (2014) was not included within this selective sweep, but it does lie within 135 kb. In a different study, 140 selective sweeps were identified from a panel of 500 soybean and wild soybean accessions (Han et al., 2016). These selective sweeps were twice as likely to contain QTL associated with domestication traits than the remaining genome. This indicates that using population genetics to identify regions of selection could be successfully used to detect QTL related to domestication.

REFERENCES

Akpertey, A., M. Belaffif, G.L. Graef, M.A. Rouf Mian, J. Grover Shannon, P.B. Cregan, M.E. Hudson, B.W. Diers, and R.L. Nelson. 2014. Effects of selective genetic introgression from wild soybean to soybean. Crop Sci. 54(6): 2683–2695.

Alpert, K.B., S. Grandillo, and S.D. Tanksley. 1995. *fw 2.2*: A major QTL controlling fruit weight is common to both red- and green-fruited tomato species. Theor. Appl. Genet. 91(6–7): 994–1000.

Anderson, J.E., T.J.Y. Kono, R.M. Stupar, M.B. Kantar, and P.L. Morrell. 2016. Environmental association analyses identify candidates for abiotic stress tolerance in *Glycine soja* , the wild progenitor of cultivated soybeans. G3:Genes|Genomes|Genetics 6(4): 835–843.

Bailey, M.A., M.A.R. Mian, T.E. Carter, D.A. Ashley, and H.R. Boerma. 1997. Pod dehiscence of soybean: Identification of quantitative trait loci. J. Hered. 199788(2): 4–6.

Cai, H.W., and H. Morishima. 2000. Genomic regions affecting seed shattering and seed dormancy in rice. Theor. Appl. Genet. 100(6): 840–846.

Caicedo, A.L., S.H. Williamson, R.D. Hernandez, A. Boyko, A. Fledel-Alon, T.L. York, N.R. Polato, K.M. Olsen, R. Nielsen, S.R. McCouch, C.D. Bustamante, and M.D. Purugganan. 2007. Genome-wide patterns of nucleotide polymorphism in domesticated rice. PLoS Genet. 3(9): 1745–1756.

Carpenter, J.A., and W.R. Fehr. 1986. Genetic variability for desirable agronomic traits in populations containing *Glycine soja* germplasm. Crop Sci. 26(4): 681.

Carter, T.E., R.L. Nelson, C.H. Sneller, and Z. Cui. 2004. Genetic diversity in soybean. p. 303–416. *In* Boerma, H.R., Specht, J.E. (eds.), Soybeans: Improvement, production, and uses. American Society of Agronomy, Madison, WI.

Chen, Y., and R.L. Nelson. 2004. Genetic variation and relationships among cultivated, wild, and semiwild soybean. Crop Sci. 44(1): 316.

Chung, W.H., N. Jeong, J. Kim, W.K. Lee, Y.G. Lee, S.H. Lee, W. Yoon, J.H. Kim, I.Y. Choi, H.K. Choi, J.K. Moon, N. Kim, and S.C. Jeong. 2014. Population structure and domestication revealed by high-depth resequencing of Korean cultivated and wild soybean genomes. DNA Res. 21(2): 153–167.

Clint Nesbitt, T., and S.D. Tanksley. 2002. Comparative sequencing in the genus Lycopersicon: Implications for the evolution of fruit size in the domestication of cultivated tomatoes. Genetics 162(1): 365–379.

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(4): 575–582.

Cong, B., J. Liu, and S.D. Tanksley. 2002. Natural alleles at a tomato fruit size quantitative trait locus differ by heterochronic regulatory mutations. Proc. Natl. Acad. Sci. 99(21): 13606–11.

Cui, Z., T.E. Carter, and J.W. Burton. 2000. Genetic base of Chinese soybean cultivars released during 1923 to 1995. Crop Sci. 40(5): 1470–1481.

Diers, B.W., P.R. Arelli, S.R. Carlson, W.R. Fehr, E.A. Kabelka, R.C. Shoemaker, and D. Wang. 2005. Registration of LDX01-1-65 soybean germplasm with soybean cyst nematode resistance derived from *Glycine soja*. Crop Sci. 45(4): 1671–1672.

Diers, B.W., P. Keim, W.R. Fehr, and R.C. Shoemaker. 1992. RFLP analysis of soybean seed protein and oil content. Theor. Appl. Genet. 83(5).

Doebley, J.F., B.S. Gaut, and B.D. Smith. 2006. The molecular genetics of crop domestication. Cell 127(7): 1309–1321.

Doebley, J., and A. Stec. 1993. Inheritance of the morphological differences between maize and teosinte: Comparison of results for two F2 populations. Genetics 134(2): 559–570.

Doebley, J., A. Stec, and C. Gustus. 1995. *teosinte branched1* and the origin of maize: Evidence for epistasis and the evolution of dominance. Genetics 141(1): 333–346.

Doebley, J., A. Stec, and L. Hubbard. 1997. The evolution of apical dominance in maize. Nature 386(6624): 485–8.

Dong, Y., X. Yang, J. Liu, B.H. Wang, B.L. Liu, and Y.Z. Wang. 2014. Pod shattering resistance associated with domestication is mediated by a NAC gene in soybean. Nat. Commun. 5: 3352.

Dorweiler, J., A. Stec, J. Kermicle, and J. Doebley. 1993. *Teosinte glume architecture 1*: A genetic locus controlling a key step in maize evolution. Science. 262(5131): 233–235.

Ebert, A.W., and R. Schafleitner. 2015. Utilization of wild relatives in the breeding of tomato and other major vegetables. p. 141–172. *In* Redden, R., Yadav, S.S., Maxted, N., Dulloo, M.E., Guarino, L., Smith, P. (eds.), Crop Wild Relatives and Climate Change. John Wiley & Sons, Inc, Hoboken, NJ, USA.

FAOSTAT. 2016. Food and Agriculture Organization of the United Nations. Available at http://www.fao.org/faostat/en/#data/QC (verified 1 January 2016).

Frary, A., T.C. Nesbitt, S. Grandillo, E. Knaap, B. Cong, J. Liu, J. Meller, R. Elber, K.B. Alpert, and S.D. Tanksley. 2000. *Fw2.2*: A quantitative trait locus key to the evolution of tomato fruit size. Science. 289(5476): 85–88.

Fridman, E., F. Carrari, Y.S. Liu, A.R. Fernie, and D. Zamir. 2004. Zooming in on a quantitative trait for tomato yield using interspecific introgressions. Science. 305(5691): 1786–1789.

Fry, W.E., and S.B. Goodwin. 1997. Re-emergence of potato and tomato late blight in the United States and Canada. Plant Dis. 81(12): 1249–1357.

Funatsuki, H., M. Suzuki, A. Hirose, H. Inaba, T. Yamada, M. Hajika, K. Komatsu, T. Katayama, T. Sayama, M. Ishimoto, and K. Fujino. 2014. Molecular basis of a shattering resistance boosting global dissemination of soybean. Proc. Natl. Acad. Sci. 111(50): 17797–17802.

Gizlice, Z., T.E. Carter, and J.W. Burton. 1994. Genetic base for North American public soybean cultivars released between 1947 and 1988. Crop Sci. 34(5): 1143–1151.

Gore, M.A., J.M. Chia, R.J. Elshire, Q. Sun, E.S. Ersoz, B.L. Hurwitz, J.A. Peiffer, M.D. McMullen, G.S. Grills, J. Ross-Ibarra, D.H. Ware, and E.S. Buckler. 2009. A first-generation haplotype map of maize. Science. 326(5956): 1115–1117.

Guo, J., Y. Wang, C. Song, J. Zhou, L. Qiu, H. Huang, and Y. Wang. 2010. A single origin and moderate bottleneck during domestication of soybean (*Glycine max*): Implications from microsatellites and nucleotide sequences. Ann. Bot. 106(3): 505–514.

Hammer, K. 1984. Das domestikationssyndrom. Die Kult. 32(3): 11–34.

Han, Y., X. Zhao, D. Liu, Y. Li, D.A. Lightfoot, Z. Yang, L. Zhao, G. Zhou, Z. Wang, L. Huang, Z. Zhang, L. Qiu, H. Zheng, and W. Li. 2016. Domestication footprints anchor genomic regions of agronomic importance in soybeans. New Phytol. 209(2): 871–884.

Harlan, J.R. 1995. The Living Fields. Cambridge University Press, Cambridge.

Harlan, J.R., and J.M.J. de Wet. 1971. Toward a rational classification of cultivated plants. Taxon 20: 509–517.

Hodgkin, T., and R. Hajjar. 2008. Using crop wild relatives for crop improvement: Trends and perspectives. p. 535–548. *In* Maxted, N., Ford-Lloyd, B. V., Kell, S.P., Iriondo, J.M., Dulloo, M.E., Turok, J. (eds.), Crop Wild Relative Conservation and Use. CABI, Wallingford, UK.

Hoisington, D., M. Khairallah, T. Reeves, J.M. Ribaut, B. Skovmand, S. Taba, and M. Warburton. 1999. Plant genetic resources: What can they contribute toward increased crop productivity? Proc. Natl. Acad. Sci. 96(11): 5937–5943.

Hufford, M.B., X. Xu, J. van Heerwaarden, T. Pyhäjärvi, J.M. Chia, R.A. Cartwright, R.J. Elshire, J.C. Glaubitz, K.E. Guill, S.M. Kaeppler, J. Lai, P.L. Morrell, L.M. Shannon, C. Song, N.M. Springer, R.A. Swanson-Wagner, P. Tiffin, J. Wang, G. Zhang, J. Doebley, M.D. McMullen, D. Ware, E.S. Buckler, S. Yang, and J. Ross-Ibarra. 2012. Comparative population genomics of maize domestication and improvement. Nat. Genet. 44(7): 808–811.

Hymowitz, T. 1970. On the domestication of the soybean. Econ. Bot. 24: 408–421.

Hymowitz, T. 1987. Introduction of the soybean to Illinois. Econ. Bot. 41(1): 28–32.

Hymowitz, T., and J.R. Harlan. 1983. Introduction of soybean to North America by Samuel Bowen in 1765. Econ. Bot. 37(4): 371–379.

Hyten, D.L., Q. Song, Y. Zhu, I.Y. 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. 103(45): 16666–16671.

Keim, P., B.W. Diers, T.C. Olson, and R.C. Shoemaker. 1990a. RFLP mapping in soybean: Association between marker loci and variation in quantitative traits. Genetics 126(3): 735–742.

Keim, P., B.W. Diers, and R.C. Shoemaker. 1990b. Genetic analysis of soybean hard seededness with molecular markers. Theor. Appl. Genet. 79(4): 465–469.

Kim, M., and B.W. Diers. 2013. Fine mapping of the SCN resistance QTL and from PI 468916. Crop Sci. 53(3): 775.

Kim, M., D.L. Hyten, T.L. Niblack, and B.W. Diers. 2011. Stacking resistance alleles from wild and domestic soybean sources improves soybean cyst nematode resistance. Crop Sci. 51(3): 934.

Konishi, S., T. Izawa, S.Y. Lin, K. Ebana, Y. Fukuta, T. Sasaki, and M. Yano. 2006. An SNP caused loss of seed shattering during rice domestication. Science. 312(5778): 1392–1396.

Kulkarni, K.P., M. Kim, J.G. Shannon, and J.-D. Lee. 2016. Identification of quantitative trait loci controlling soybean seed weight in recombinant inbred lines derived from PI 483463 (*Glycine soja*) × "Hutcheson" (*G. max*). Plant Breed. 135(5): 614–620.

Ladizinsky, G., C.A. Newell, and T. Hymowitz. 1979. Wide crosses in soybeans: Prospects and limitations. Euphytica 28(2): 421–423.

Lam, H.M., X. Xu, X. Liu, W. Chen, G. Yang, F.L. Wong, M.W. Li, W. He, N. Qin, B. Wang, J. Li, M. Jian, J.J. Wang, G. Shao, J.J. Wang, S.S.M. Sun, and G. Zhang. 2010. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. Nat. Genet. 42(12): 1053–1059.

Lee, G.A., G.W. Crawford, L. Liu, Y. Sasaki, and X. Chen. 2011. Archaeological soybean (*Glycine max*) in East Asia: Does size matter? PLoS One 6(11): e26720.

Li, Y., R. Guan, Z. Liu, Y. Ma, L. Wang, L. Li, F. Lin, W. Luan, P. Chen, Z. Yan, Y. Guan, L. Zhu, X. Ning, M.J.M. Smulders, W. Li, R. Piao, Y. Cui, Z. Yu, M. Guan, R. Chang, A. Hou, A. Shi, B. Zhang, S. Zhu, and L. Qiu. 2008a. Genetic structure and diversity of cultivated soybean (*Glycine max* (L.) Merr.) landraces in China. Theor. Appl. Genet. 117(6): 857–871.

Li, Y.H., W. Li, C. Zhang, L. Yang, R.Z. Chang, B.S. Gaut, and L.J. Qiu. 2010. Genetic diversity in domesticated soybean (*Glycine max*) and its wild progenitor (*Glycine soja*) for simple sequence repeat and single-nucleotide polymorphism loci. New Phytol. 188(1): 242–253.

Li, D., T.W. Pfeiffer, and P.L. Cornelius. 2008b. Soybean QTL for yield and yield components associated with *Glycine soja* alleles. Crop Sci. 48(2): 571–581.

Li, Y., S. Zhao, J. Ma, D. Li, L. Yan, J. Li, X. Qi, X. Guo, L. Zhang, W. He, R. Chang, Q. Liang, Y. Guo, C. Ye, X. Wang, Y. Tao, R. Guan, J. Wang, Y. Liu, L. Jin, X. Zhang, Z. Liu, L. Zhang, J. Chen, K. Wang, R. Nielsen, R. Li, P. Chen, W. Li, J.C. Reif, M. Purugganan, J. Wang, M. Zhang, J. Wang, and L. Qiu. 2013. Molecular footprints of domestication and improvement in soybean revealed by whole genome re-sequencing. BMC Genomics 14(1): 579.

Li, Y.H., G. Zhou, J. Ma, W. Jiang, L.G. Jin, Z. Zhang, Y. Guo, J. Zhang, Y. Sui, L. Zheng, S.S. Zhang, Q. Zuo, X.H. Shi, Y.F. Li, W.K. Zhang, Y. Hu, G. Kong, H.L. Hong, B. Tan, J. Song, Z.X. Liu, Y. Wang, H. Ruan, C.K.L. Yeung, J. Liu, H. Wang, L.J. Zhang, R.X. Guan, K.J. Wang, W.B. Li, S.Y. Chen, R.Z. Chang, Z. Jiang, S.A. Jackson, R. Li, and L.J. Qiu. 2014. De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. Nat. Biotechnol. 32(10): 1045–1052.

Li, C., A. Zhou, and T. Sang. 2006a. Genetic analysis of rice domestication syndrome with the wild annual species, *Oryza nivara*. New Phytol. 170(1): 185–194.

Li, C., A. Zhou, and T. Sang. 2006b. Rice domestication by reducing shattering. Science. 311(5769): 1936–1939.

Lin, S.Y., T. Sasaki, and M. Yano. 1998. Mapping quantitative trait loci controlling seed dormancy and heading date in rice, *Oryza sativa L.*, using backcross inbred lines. Theor Appl Genet 96: 997–1003.

Liu, B., T. Fujita, Z.H. Yan, S. Sakamoto, D. Xu, and J. Abe. 2007. QTL mapping of domestication-related traits in soybean (*Glycine max*). Ann. Bot. 100(5): 1027–1038.

Liu, B., S. Watanabe, T. Uchiyama, F. Kong, A. Kanazawa, Z. Xia, A. Nagamatsu, M. Arai, T. Yamada, K. Kitamura, C. Masuta, K. Harada, and J. Abe. 2010. The soybean stem growth habit gene Dt1 is an ortholog of Arabidopsis TERMINAL FLOWER1. Plant Physiol. 153(1): 198–210.

Luo, X., X. Bai, X. Sun, D. Zhu, B. Liu, W. Ji, H. Cai, L. Cao, J. Wu, M. Hu, X. Liu, L. Tang, and Y. Zhu. 2013. Expression of wild soybean WRKY20 in Arabidopsis enhances drought tolerance and regulates ABA signalling. J. Exp. Bot. 64(8): 2155–2169.

Maxted, N., and S. Kell. 2009. Establishment of a network for the in situ conservation of crop wild relatives: Status and needs. Food and Agriculture Organization of the United Nations, Rome.

Mikel, M.A. 2008. Genetic diversity and improvement of contemporary proprietary north American dent corn. Crop Sci. 48(5): 1686–1695.

Mikel, M.A., B.W. Diers, R.L. Nelson, and H.H. Smith. 2010. Genetic diversity and agronomic improvement of north american soybean germplasm. Crop Sci. 50(4): 1220–1228.

Niblack, T.L., K.N. Lambert, and G.L. Tylka. 2006. A model plant pathogen from the kingdom Animalia: *Heterodera glycines*, the soybean cyst nematode. Annu. Rev. Phytopathol. 44(1): 283–303.

Nichols, D.M., K.D. Glover, S.R. Carlson, J.E. Specht, and B.W. Diers. 2006. Fine mapping of a seed protein QTL on soybean linkage group I and its correlated effects on agronomic traits. Crop Sci. 46(2): 834–839.

Placido, D.F., M.T. Campbell, J.J. Folsom, X. Cui, G.R. Kruger, P.S. Baenziger, and H. Walia. 2013. Introgression of novel traits from a wild wheat relative improves drought adaptation in wheat. Plant Physiol. 161(4): 1806–19.

Qi, X., M.W. Li, M. Xie, X. Liu, M. Ni, G. Shao, C. Song, Y.A. Kay-Yuen, Y. Tao, F.L. Wong, S. Isobe, C.F. Wong, K. Sen Wong, C. Xu, C. Li, Y. Wang, R. Guan, F. Sun, G. Fan, Z. Xiao, F. Zhou, T.H. Phang, X. Liu, S.W. Tong, T.F. Chan, S.M. Yiu, S. Tabata, J. Wang, X. Xu, and H.M. Lam. 2014. Identification of a novel salt tolerance gene in wild soybean by whole-genome sequencing. Nat. Commun. 5: 4340.

Ram, T., N.D. Majumder, D. Krishnaveni, and M.M. Ansari. 2007. Rice variety Dhanrasi, an example of improving yield potential and disease resistance by introgressing gene(s) from wild species (*Oryza rufipogon*). Curr. Sci. 92(7): 987–992.

Redden, R., S.S. Yadav, N. Maxted, M.E. Dulloo, L. Guarino, and P. Smith (Eds). 2015. Crop wild relatives and climate change. Wiley-Blackwell, Hoboken, NJ.

Rick, C.M. 1974. High soluble-solids content in large-fruited tomato lines derived from a wild green-fruited species. Hilgardia 42(15): 493–510.

Ross-Ibarra, J., P.L. Morrell, and B.S. Gaut. 2007. Plant domestication, a unique opportunity to identify the genetic basis of adaptation. Proc. Natl. Acad. Sci. 104(Supplement 1): 8641–8648.

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(5): 1438.

Sherman-Broyles, S., A. Bombarely, A.F. Powell, J.L. Doyle, A.N. Egan, J.E. Coate, and J.J. Doyle. 2014. The wild side of a major crop: Soybean's perennial cousins from Down Under. Am. J. Bot. 101(10): 1651–1665.

Singh, R.J., and R.L. Nelson. 2015. Intersubgeneric hybridization between *Glycine max* and *G. tomentella*: production of F1, amphidiploid, BC1, BC2, BC3, and fertile soybean plants. Theor. Appl. Genet. 128(6): 1117–1136.

Sugimoto, K., Y. Takeuchi, K. Ebana, A. Miyao, H. Hirochika, N. Hara, K. Ishiyama, M. Kobayashi, Y. Ban, T. Hattori, and M. Yano. 2010. Molecular cloning of *Sdr4*, a regulator involved in seed dormancy and domestication of rice. Proc. Natl. Acad. Sci. 107(13): 5792–5797.

Sun, L., Z. Miao, C. Cai, D. Zhang, M. Zhao, Y. Wu, X. Zhang, S.A. Swarm, L. Zhou, Z.J. Zhang, R.L. Nelson, and J. Ma. 2015. *GmHs1-1*, encoding a calcineurin-like protein, controls hard-seededness in soybean. Nat. Genet. 47(8).

Suzuki, M., K. Fujino, and H. Funatsuki. 2009. A major soybean QTL, *qPDH1*, controls pod dehiscence without marked morphological change. Plant Prod. Sci. 12(2): 217–223.

Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics 105(2): 437–460.

Tang, L., H. Cai, W. Ji, X. Luo, Z. Wang, J. Wu, X. Wang, L. Cui, Y. Wang, Y. Zhu, and X. Bai. 2013. Overexpression of *GsZFP1* enhances salt and drought tolerance in transgenic alfalfa (*Medicago sativa* L.). Plant Physiol. Biochem. 71: 22–30.

Tian, F., D.J. Li, Q. Fu, Z.F. Zhu, Y.C. Fu, X.K. Wang, and C.Q. Sun. 2006. Construction of introgression lines carrying wild rice (*Oryza rufipogon* Griff.) segments in cultivated rice (*Oryza sativa* L.) background and characterization of introgressed segments associated with yield-related traits. Theor. Appl. Genet. 112(3): 570–580.

Tian, Z.X., X.B. Wang, R. Lee, Y.H. Li, J.E. Specht, R.L. Nelson, P.E. McClean, L.J. Qiu, and J.X. Ma. 2010. Artificial selection for determinate growth habit in soybean. Proc. Natl. Acad. Sci. 107(19): 8563–8568.

USDA. 2016. United States Department of Agriculture (USDA) Foreign Agricultural Service's Database. Available at http://www.fas.usda.gov/data/world-agricultural-production.

Van, K., M.Y. Kim, J.H. Shin, D.K. Kyung, Y.H. Lee, and S.H. Lee. 2014. Molecular evidence for soybean domestication. p. 465–481. *In* Tuberosa, R., Graner, A., Frison, E. (eds.), Genomics of Plant Genetic Resources: Volume 1. Managing, Sequencing and Mining Genetic Resources.

Walsh, B. 2008. Using molecular markers for detecting domestication, improvement, and adaptation genes. Euphytica 161(1–2): 1–17.

Wang, D., P.R. Arelli, R.C. Shoemaker, and B.W. Diers. 2001. Loci underlying resistance to Race 3 of soybean cyst nematode in *Glycine soja* plant introduction 468916. Theor. Appl. Genet. 103(3): 561–566.

Wang, J., S. Chu, H. Zhang, Y. Zhu, H. Cheng, and D. Yu. 2016a. Development and application of a novel genome-wide SNP array reveals domestication history in soybean. Sci. Rep. 6: 20728.

Wang, L., R. Guan, L. Zhangxiong, R. Chang, and L. Qiu. 2006. Genetic diversity of Chinese cultivated soybean revealed by SSR markers. Crop Sci. 46(3): 1032–1038.

Wang, W., Q. He, H. Yang, S. Xiang, T. Zhao, and J. Gai. 2013. Development of a chromosome segment substitution line population with wild soybean (*Glycine soja* Sieb. et Zucc.) as donor parent. Euphytica 189(2): 293–307.

Wang, L., F. Lin, L. Li, W. Li, Z. Yan, W. Luan, R. Piao, Y. Guan, X. Ning, L. Zhu, Y. Ma, Z. Dong, H. Zhang, Y. Zhang, R. Guan, Y. Li, Z. Liu, R. Chang, and L. Qiu. 2016b. Genetic diversity center of cultivated soybean (*Glycine max*) in China – New insight and evidence for the diversity center of Chinese cultivated soybean. J. Integr. Agric. 15(11): 2481–2487.

Wang, S., and Z. Lu. 2006. Genetic diversity among parental lines of Indica hybrid rice (*Oryza sativa* L.) in China based on coefficient of parentage. Plant Breed. 125(6): 606–612.

Wang, H., T. Nussbaum-Wagler, B. Li, Q. Zhao, Y. Vigouroux, M. Faller, K. Bomblies-Yant, L. Lukens, and J. Doebley. 2005. The origin of the naked grains of maize. Nature 436(7051): 714–719.

Watterson, G.A. 1975. On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. 7(2): 256–276.

Woodworth, C.M. 1932. Genetics and breeding in the improvement of the soybean. Illinois Agr Exp Sta Bull 384: 297–404.

Xin, D., Z. Qi, H. Jiang, Z. Hu, R. Zhu, J. Hu, H. Han, G. Hu, C. Liu, and Q. Chen. 2016. QTL location and epistatic effect analysis of 100-seed weight using wild soybean (*Glycine soja* Sieb. & Zucc.) chromosome segment substitution lines. PLoS One 11(3): e0149380.

Xiong, L.X., K.D. Liu, X.K. Dai, C.G. Xu, and Q.F. Zhang. 1999. Identification of genetic factors controlling domestication-related traits of rice using an F-2 population of a cross between *Oryza sativa* and *O. rufipogon*. Theor. Appl. Genet. 98(2): 243–251.

Xu, D.H., J. Abe, J.Y. Gai, and Y. Shimamoto. 2002. Diversity of chloroplast DNA SSRs in wild and cultivated soybeans: Evidence for multiple origins of cultivated soybean. Theor. Appl. Genet. 105(5): 645–653.

Yang, L., K. Wu, P. Gao, X. Liu, G. Li, and Z. Wu. 2014. GsLRPK, a novel cold-activated leucine-rich repeat receptor-like protein kinase from *Glycine soja*, is a positive regulator to cold stress tolerance. Plant Sci. 215–216: 19–28.

Zhang, C., L. Liu, X. Wang, J. Vossen, G. Li, T. Li, Z. Zheng, J. Gao, Y. Guo, R.G.F. Visser, J. Li, Y. Bai, and Y. Du. 2014. The *Ph-3* gene from *Solanum pimpinellifolium* encodes CC-NBS-LRR protein conferring resistance to *Phytophthora infestans*. Theor. Appl. Genet. 127(6): 1353–1364.

Zhang, H., N. Mittal, L.J. Leamy, O. Barazani, and B.H. Song. 2017. Back into the wild-Apply untapped genetic diversity of wild relatives for crop improvement. Evol. Appl. 10(1): 5–24.

Zhou, Z., Y. Jiang, Z. Wang, Z. Gou, J. Lyu, W. Li, Y. Yu, L. Shu, Y. Zhao, Y. Ma, C. Fang, Y. Shen, T. Liu, C. Li, Q. Li, M. Wu, M. Wang, Y. Wu, Y. Dong, W. Wan, X. Wang, Z. Ding, Y. Gao, H. Xiang, B. Zhu, S.H. Lee, W. Wang, and Z. Tian. 2015. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. Nat. Biotechnol. 33(4): 408–14.

Zhou, L., L. Luo, J.F. Zuo, L. Yang, L. Zhang, X. Guang, Y. Niu, J. Jian, Q.C. Geng, L. Liang, Q. Song, J.M. Dunwell, Z. Wu, J. Wen, Y.Q. Liu, and Y.M. Zhang. 2016. Identification and validation of candidate genes associated with domesticated and improved traits in soybean. Plant Genome 9(2).

Zhu, D., H. Cai, X. Luo, X. Bai, M.K. Deyholos, Q. Chen, C. Chen, W. Ji, and Y. Zhu. 2012. Over-expression of a novel JAZ family gene from *Glycine soja*, increases salt and alkali stress tolerance. Biochem. Biophys. Res. Commun. 426(2): 273–279

CHAPTER TWO

QTL Mapping of Domestication-Related Traits

INTRODUCTION

The domestication of crop species from wild progenitors led to a common pattern in morphological changes, often referred to as the 'domestication syndrome' (Hammer, 1984). Among the traits typically associated with domestication are a loss of seed dispersal, increased apical dominance, larger grains, and more robust plant growth (Doebley et al., 2006). Domestication-related traits (DRT) arose through human selection due to their importance for cultivation in agriculture systems. Larger seeds and reduced shattering led to higher yields, and morphological changes that increase upright growth facilitated harvesting. Discovering the genes controlling these important agronomic traits helps us to understand the process of domestication. These genes are also commonly located in regions of low genetic diversity, most likely due to the strong selection pressure imposed during domestication (Zhou et al., 2015).

Research has been conducted across numerous crops to identify genes or QTL controlling DRT. Two genes important for the development of maize were *tb1* (Doebley and Stec, 1993; Doebley et al., 1995) and *tga1*(Dorweiler et al., 1993). These genes were responsible for the change in branching architecture (*tb1*) and the loss of the hard casing surrounding the kernel of the wild ancestor (*tga1*). Some other economically important crops that have been researched for DRT are rice (Li et al., 2006b; Sugimoto et al., 2010), wheat (Peng et al., 2003; Simons, 2005), and tomato (Frary et al., 2000). However, relatively few studies have been conducted to identify QTL for DRT in soybean, and only three genes related to domestication have been cloned (Tian et al., 2010; Dong et al., 2014; Sun et al., 2015).

Soybean was most likely domesticated from the wild annual species, *Glycine soja* (Seib. and Zucc.), 3,000-5,000 years ago in East Asia (Carter et al., 2004). Compared to the wild progenitor, cultivated soybean exhibits upright growth dominated by a single mainstem, large yellow seeds, and a general lack of pod dehiscence (shattering). Wild soybean, on the other hand, displays a procumbent or climbing growth habit, with small black seeds, dehiscent pods, and prolific branching (Broich and Palmer, 1980). Despite the difference in morphological appearance, *G. max* and *G. soja* share a high genomic similarity and stable chromosomal pairing allows fertile hybridization (Singh and Hymowitz, 1988).

A few studies have been conducted with the objective of mapping QTL for soybean DRT (Keim et al., 1990a; Liu et al., 2007; Wang et al., 2013; Kulkarni et al., 2016). Many studies using *G. soja*-derived populations focus on yield related traits, such as 100-seed weight (Maughan et al., 1996; Xin et al., 2016; Kulkarni et al., 2016), while only a few studies report a comprehensive analysis of morphological DRT (Liu et al., 2007). The studies reported thus far are also typically limited in population size by using fewer than 200 individuals. While smaller populations are more conducive for genotyping and collecting phenotypic data, larger numbers of individuals increase the power of the mapping experiment and permit the detection of more QTL (Beavis, 1998). To this end, the objective of this experiment was to map QTL associated with soybean DRT using a larger number of recombinant-inbred lines (RIL) in order to identify additional QTL that may have been undetected in previous studies. These QTL would be of importance to better understand the basis of agronomic traits in soybean and to identify regions that would have been subjected to selection during domestication.

MATERIALS AND METHODS

Plant Materials

To generate RIL populations, crosses were made between the *G. max* line 'Williams 82' and the *G. soja* accessions PI 468916 (a maturity group III accession from Liaoning Province, China) and PI 479752 (a maturity group I accession from Jilin Province, China). Both *G. soja* accessions were used to develop the SoySNP50K Illumina Infinium BeadChip (Song et al., 2013), and had been previously used to construct interspecific populations (Keim et al., 1990a; Song et al., 2016). The F_2 progeny were advanced through single-seed descent to the F_6 generation. Due to an error in accession identification before crossing, the particular *G. soja* parent for each RIL was not known and therefore molecular markers were later used to discriminate the two populations. In 2012, 3014 F_6 lines were grown at Urbana, IL in 1.2 m plots with 0.76 m between rows and 1.1 m alleyways. Eight traits (maturity, main stem length, lodging, stem diameter, growth habit, leaflet length, leaflet shape, and 100-seed weight) were measured in 2012 and used to select lines for replication in 2013-2015. A total of 800 lines were chosen to retain the maximum and minimum for each trait and for all pairs of traits, while also representing the full distribution between the extremes.

Trait Evaluation

Traits that were measured included flowering date (R1), defined as the date when 50% of the plants produced at least one flower; plant maturity (R8), defined as the date when 95% of the pods in a plot reached their mature color (Fehr et al., 1971); main stem length (cm), measured from the soil surface to the top node of the main stem; lodging scored at maturity on a scale of 1 to 9 (1 = erect plants 90° from soil surface, 9 = completely lodged at 0° from soil surface); stem diameter (mm), averaged from three plants measured with digital calipers below the univariate node; growth habit, scored on a 1 to 5 scale with 0.5 increments (1 = indeterminate soybean growth type, 5 = vining growth similar to *G. soja*); leaflet length, scored on a 1 to 5 scale with 0.5 increments ($1 \leq 4$ cm, $5 \geq 12$ cm); leaflet shape, classified as broad, narrow, or intermediate; shattering was rated at maturity (R8) and 10 days after maturity using a 1 to 5 scale based on the estimated percentage of shattered pods (1 = 0%, 2 = 1-10%, 3 = 11-25%, 4 = 26-50%, 5 = 51-80%, 6 = 81-100%); pubescence form was classified as erect, semi-appressed, and appressed as described by Bernard (1975) and Lee et al. (1999), with the addition of a very-appressed classification which exhibited pubescence completely flat on the top of the leaf; and 100-seed weight (g).

All eleven traits were measured at Urbana, IL environments with the exception of pubescence form, which was only measured in 2014 and 2015. At West Lafayette, IN, stem diameter was not measured; shattering, pubescence form, growth habit, and R1 were measured in 2015 only; and leaf length and shape were measured in 2013 and 2015.

Experimental Design and Statistical Analysis

The 800 RILs were grown in a randomized block design with two replications. The experiment was conducted at two locations (Urbana, IL and West Lafayette, IN) during 2013, 2014, and 2015. Seeds were sown at a rate of 33 seeds per meter in plots 1.2 m in length with 0.76 m alleyways. RILs with main stem lengths over 100 cm in 2012 were grown with 1.52 m between plots to reduce competition between rows, while the other RILs were grown with 0.76 m spacing.

An analysis of variance was conducted in R (R Core Team, 2014) using the ‘lme4’ package for mixed models (Bates et al., 2015). The genotypes were considered as fixed effects, while environments (year by location), replications within environments, and genotype by

environment were treated as random effects. Least-squares means of the genotypes were calculated using the 'lsmeans' package in R (Lenth, 2016). Traits whose errors were not normally distributed were transformed using the Box-Cox transformation.

DNA Isolation and Genotyping-by-Sequencing

Ten individuals from each RIL were grown in a greenhouse for tissue sampling. Newly-expanded trifoliate leaves were sampled and combined for DNA extraction. Genomic DNA was isolated using a CTAB method modified from Mace et al. (2003) and DNA samples were quantified using PicoGreen (Invitrogen, NY, USA). GBS libraries were constructed following a two enzyme protocol modified from Thurber et al. (2013). Approximately 250 ng of DNA from each RIL were digested with *HindIII* and *BfaI* and were ligated with one of 364 unique barcoded adapters with a *HindIII* overhang. This enzyme combination had previously been shown to successfully digest soybean DNA (Akpertey, 2015). The samples were then pooled and submitted to the W. M. Keck Center at the University of Illinois for single-end 100-bp sequencing with Illumina HiSeq2000. Five libraries, each consisting of 190 samples, were created and sequenced on separate Illumina lanes which averaged 180 million reads (1.1 million reads per sample). Each parent was replicated six times to provide deeper coverage compared to the RILs.

The GBSv2 pipeline in TASSEL 5.0 was used to process the sequencing reads (Glaubitz et al., 2014). Default settings were used with the exceptions of increasing the maximum tag length to 80 bp (-kmerLength 80 in the GBSSeqToTagDBPlugin), and SNPs with $F < 0.7$, minor allele frequency (MAF) < 0.01, or caused by indels were removed (via –minPosQS in ProductionSNPCallerPluginV2f). Sequenced tags were aligned to the Williams 82 reference genome (Wm82.a2.v1 downloaded from Phytozome) using Bowtie 2 with the "very-sensitive" option (Langmead et al., 2009). TASSEL was also used to filter SNPs and impute missing data. Initial SNP calling yielded 110,419 SNPs, and after removing markers with greater than 25% missing data, and over 11% heterozygous calls, 35,303 SNPs remained.

To distinguish the two *G. max* x *G. soja* populations, the first two principal components of the 35,303 SNPs were plotted. The genotypes predominantly separated into two groups associated with the two *G. soja* accessions. While the majority of the two populations could be distinguished based on the principal components, some genotypes were unable to be classified with high confidence due to poor separation through the principal components. To distinguish

these unclassified genotypes, 70 lines were chosen for screening with SSR markers, along with two samples of each of the three parents, and 10 randomly selected genotypes from each principal component group to validate the accuracy of the PC clustering. Five SSR markers were chosen that showed polymorphism between all three parents and were located on separate linkage groups. The classification of the 20 controls was confirmed based on the SSR markers, and 32 of the RILs were successfully classified into one of the two populations. With the combination of SNP and SSR markers, the Williams 82 x PI 468916 population (abbreviated as PI468) contained 170 RILs, and the Williams 82 x PI 479752 population (PI479) contained 590 RILs. A total of 53 lines were unable to be classified, likely due to poor DNA quality.

The full-sib family haplotype imputation method (FSFHap) was used in TASSEL to impute missing data and identify parental haplotypes within the PI 479752 and PI 468916 populations (Swarts et al., 2014). A total of 6 lines with less than 20% call rate for SNPs were removed prior to imputation. A total of 17,153 SNPs were polymorphic in the Williams 82 x PI 468916 population, 17,112 were polymorphic in Williams 82 x PI 479752, with 7,721 segregating in both populations. While the FSFHap method efficiently fills in missing data and identifies parental haplotypes, it also can inflate heterozygosity. After imputation, lines with heterozygosity greater than 30% were removed from the analysis. This left 151 RILs in the Williams 82 x PI 468916 population, and 510 RILs in Williams 82 x PI 479752.

QTL Mapping

A genetic map was built using the R/qtl package to verify that SNPs were correctly aligned on the soybean reference genome (Broman et al., 2003). SNPs that caused large gaps in the genetic map were removed. Plots of physical position versus genetic location (Fig. 2.3) were consistent with results from another study using a Williams 82 x PI 479752 population (Song et al., 2016) and confirmed that the markers were correctly ordered.

The final SNP set used for QTL mapping consisted of 16,893 SNPs polymorphic in Williams 82 x PI 468916, 17,052 SNPs polymorphic in Williams 82 x PI 479752, and 7,673 SNPs polymorphic in both populations. Haley-Knott regression was used in R through the lm function and least-squares means for each RIL were regressed onto the genotype. Genotypes were coded as the dosage of *G. soja* alleles (loci homozygous for *G. soja* alleles were 2 and loci with both alleles from Williams 82 as 0). Forward regression was employed by adding the most significant SNP from the previous scan as a covariate in the model. When mapping with both

populations combined, the family was included as a covariate. Significance thresholds were set using a Bonferroni correction at $\alpha = 0.01$. PhenoGram was used to display the locations of QTL on the physical map (Wolfe et al., 2013)

RESULTS AND DISCUSSION

Phenotypic Distributions

While all traits could be used to differentiate the *G. soja* accessions from Williams 82 (as shown in Table 2.1), not all should be considered domestication-related traits. Of these traits, those that were most likely selected during domestication were larger stem diameter, upright growth (reduced lodging), reduced shattering, larger 100-seed weight, earlier flowering, and altered growth habit. Table 2.2 gives the variance components (expressed as percentage) for each trait. In all cases, the genotype (RIL) explains the majority of the phenotypic variance, ranging from 92.23% for R1 (days after planting) to 50.1% for shattering. The environment explained only a minor portion of the variation, with the exception of shattering (18.4%) and R1 flowering date (22.5%). These traits are the most influenced by environment, specifically temperature and humidity in the case of shattering, and planting date in the case of flowering. When the flowering date is expressed as the number of days after planting (R1 DAP), the environmental variation largely disappears (3.6%) and the genotypic variation increases (92.2%). The interaction of genotype by environment also explained only a small portion of the variation, ranging between 14.6% (shattering) and 1.9% (R1 DAP). Based on the variance components, the majority of traits were stable across environments, with the most variability between environments seen with shattering and lodging.

The distributions of traits are given in Figure 2.1. Traits without transgressive segregation were stem diameter, shattering, growth habit, lodging, 100 seed weight, and leaflet size. Of these, stem diameter, lodging, and seed weight were skewed towards the *G. soja* parents in both populations (and shattering in PI468). This pattern in distribution has been previously observed in other *G. max* x *G. soja* populations (Cicek et al., 2006; Liu et al., 2007; Kulkarni et al., 2016) and reflects the difficulty in using *G. soja* for breeding, where multiple backcrosses are typically used to recover individuals that resemble the *G. max* parent (Carpenter and Fehr, 1986). RILs were identified that were transgressive for flowering date, maturity date, and main stem length

(Fig. 2.1). Flowering date appeared to display a bimodal distribution for both RIL populations, even when PI 479752 and Williams 82 flowered only two days apart. This bimodal distribution would suggest the involvement of one or a few major genes.

Many traits were correlated, as shown in Figure 2.2. Some of the strongest correlations were to be expected, such as between R8 and R1DAP (0.80), stem diameter (0.65), and main stem length (0.69). Main stem length and growth habit were also correlated (0.65), which is unsurprising given that vining plants tended to have prolific growth. Surprisingly, the strongest correlation was between stem diameter and leaflet size (0.82)—two seemingly unrelated characteristics. Other traits showed only weak or negligible correlations. Lodging was weakly correlated with main stem length (0.26) and stem diameter (-0.40). Both traits would seemingly affect lodging, but potentially the correlation in the set of RILs was diluted due to our selection procedure. By maintaining the pairwise combinations of extreme traits, we were able to preserve the less common genotypes (e.g. short and lodged, and small stem diameter and erect). Reducing the phenotypic correlation between traits should, in theory, allow a more efficient dissection of QTL for each of the traits.

QTL Models

QTL mapping with Haley-Knott regression resulted in the models given in Table 2.3. The number of QTL identified for each trait was dependent on the trait and the population used for mapping. The fewest QTL were identified in PI468, which contained the fewest RILs (151). The smaller population size resulted in a reduced power for QTL detection, resulting in only 1-3 loci being detected for each trait. More QTL were identified in the larger PI 479752 population (510 RILs), with between 2-10 loci per trait. The trend generally continued when both populations were used for mapping (661 RILs), resulting in 3-14 QTL. This advantage in using a multiparental mapping approach is consistent with the results found in a study using sorghum (Higgins et al., 2014). In addition to detecting more significant QTL, the QTL models built from the larger sets of RILs also explained a larger portion of the phenotypic variation as described by R^2 (Table 2.3). On average, the QTL detected from the smaller PI 468916 population collectively explained 43% of the variation, while the average R^2 for the QTL models from PI479 was 65%. Mapping with both families decreased the R^2 of the QTL model for most traits, but the changes were only marginal (largest change was -8% for leaf shape).

For some traits (R1DAP, R8, stem diameter, shattering, lodging, leaflet size), the QTL detected in PI468 were a subset of those detected in PI479 (Table 2.3, Table 2.4). Other traits (main stem length, growth habit, seed weight, leaflet size, and pubescence form) were controlled by a combination of common and unique QTL. More population specific QTL were identified in PI479, most likely due to the higher statistical power with more RILs. Unique QTL were identified in PI468 for main stem length, growth habit, seed weight, leaflet shape and pubescence form. By mapping with both populations, more QTL were identified, often capturing the unique QTL from both populations (Table 2.4). In many cases, QTL were only detected when the two populations were combined, which may be due to the increase in statistical power as more RILs are used.

Flowering Time and Maturity

Flowering time was expressed as the days after planting (R1DAP) due to the reduced environmental effect compared to the R1 measurement (Table 2.2). Tables 2.5 and 2.6 give the locations and effects of the QTLs associated with flowering time. A total of five QTL were significant in PI479, while only one QTL was detected in PI468. *qR1DAP-6* displayed the strongest effect and was detected in both populations on chr 6, and corresponded to the location of the E1 maturity gene (Xia et al., 2012). Williams 82 carries a recessive allele for this gene (*e1-as*) which leads to earlier flowering compared to *E1*. The significant association of SNPs near the *E1* locus would indicate that both *G. soja* accessions carry the functional *E1* allele. This result is somewhat unexpected given PI 479752 flowered only 2 days later than Williams 82 and both accessions matured earlier than Williams 82, which has the *e1* allele. *qR1DAP-16* was only detected in PI479. The position of the most significant marker at 4.2 Mb on chr 16 corresponds to an ortholog to the *Arabidopsis FT* gene, identified as *GmFT5a* (Kong et al., 2010). In our QTL analysis, the *G. soja* allele at this locus carried an additive effect of -3.2 days. In the Kong et al. (2010) study, ectopic expression of *GmFT5a* in *Arabidopsis* led to an early induction of flowering, which suggests that the *G. soja* allele at this locus increases the expression of *GmFT5a* relative to Williams 82. *qR1DAP-6* and *qR1DAP-16* displayed the strongest interaction among QTL, explaining approximately 5% of the variation (Table 2.7). Epistatic interactions account for only a small proportion of variance compared to the additive effects of the QTL.

A major QTL for maturity (*qR8-6*) was also detected on chr 6 in the same location as *qR1DAP-6* (Table 2.4). This is due to the fact that *E1* affects both the time of flowering and

maturity. The majority of QTL controlling flowering time were also significantly associated with maturity (Table 2.4). Of the nine R8 QTL, only three (*qR8-13*, *qR8-19*, and *qR8-20*) did not overlap with QTL for flowering time (Table 2.6). The *G. soja* alleles at these three QTL all contributed negative effects (earlier maturity), which is in the direction of the *G. soja* phenotypes relative to Williams 82. The common genetic basis for the two traits explains the strong correlation (0.8) between the traits.

Shattering

Shattering is one of the few soybean domestication-related traits for which major genes have been cloned. Dong et al. (2014) used a candidate gene approach to identify the *SHAT1-5* gene as *Glyma16g02200*, whose overexpression is associated with thickened cell walls in the ventral suture and reduced shattering. Funatsuki et al. (2014) used near isogenic lines derived from a cross between two *G. max* cultivars to fine map a separate *PDH1* gene (*Glyma16g25580*) encoding a dirigent-like protein. While *SHAT1-5* influences the binding forces of the pod suture, *PDH1* is believed to affect the torsion forces of the pod. Using the *G. max* x *G. soja* RILs in this study, shattering was largely controlled by a single QTL on chr 16 (*qSH-16*) which explained 23%-53% of the variation (Table 2.4). The position of *qSH-16* at 30 Mb on chr 16 correlates closely with *PDH1* at 29.6 Mb (Funatsuki et al., 2014). This seems to indicate that *PDH1* plays a more important role than *SHAT1-5* in these two *G. max* x *G. soja* populations.

While major genes controlling shattering have been cloned, the trait is still quantitative in nature. In addition to *qSH-16*, several QTL with smaller effects were identified (Tables 2.4 and 2.6). *qSH-16* was the only QTL identified in PI468, but five additional QTL were detected in PI479. One of these, *qSH-6*, is likely due to *E1* and was the only QTL for which the *G. soja* allele carried a negative effect (decreasing shattering). As shown in a review of shattering QTL (Funatsuki et al., 2012), *qSH-2*, *qSH-15*, and *qSH-19* were also detected in other populations. However, these minor QTL are detected with less consistency than the QTL on chr 16 and may therefore represent population specific QTL, such as *qSH-7*, which was not detected among the six studies surveyed by Funatsuki et al. (2012). A significant interaction was detected between *qSH-16* and *qSH-15* (Table 2.7).

Stem Diameter

Two QTL were identified in PI468 and together explained 28% of the variance. In addition to these QTL, eight more were identified in the larger PI 479752 population, and a much larger proportion of the phenotypic variance (62%) was explained by the QTL model. All QTL related to stem diameter had only a small effect and individually explained modest proportions of the variance (Tables 2.5 and 2.6). The QTL for this trait had additive effects of up to -0.3 mm for the substitution of a *G. soja* allele. As expected, the effects of the *G. soja* alleles were in the direction of the *G. soja* parent, with the exception of *qSD-6* in PI479. This QTL also appears to be controlled by the *E1* locus and reflects the positive correlation between stem diameter and maturity. While both QTL (*qSD-13, qSD-1*) from PI468 were also detected in PI479, they played a smaller role. Interestingly, the three QTL with the largest phenotypic variation explained (PVE) in PI479 (*qSD-19, qSD-6,* and *qSD-12*) were not detected in PI468. This may indicate that *G. soja* accessions carry different QTL for stem diameter. Chen et al. (2011) used a biparental soybean population to map several traits related to stem strengths, including stem diameter. Four QTL from PI479 could also be found in this soybean population, namely *qSD-6, qSD-7, qSD-14*, and *qSD-19*. Therefore, these four loci exhibit variation within *G. max* and likely were not subjected to fixation during domestication. Using another *G. max* x PI 468916 population, Keim et al. (1990) also mapped QTL associated with stem diameter. Three linked markers on chr 19 (LG L) were significant in that study, but interestingly were absent from PI468 in this current study. However, *qSD-19* carried the largest PVE in PI479 and was located in the same region as the QTL found by Keim et al. (1990).

Main Stem Length

As would be expected, the *E1* locus was significantly associated with main stem length as the *qMSL-6* QTL, and explained a sizable amount of the variation with a PVE of 29% and 32% in the PI 468916 and PI 479752 populations respectively (Table 2.4). This was the only QTL shared between the two individual populations, although additional QTL specific to each population were also detected with large PVE (*qMSL-1* and *qMSL-13* in PI468 and *qMSL-12* in PI479). *qMSL-6* is the only QTL for which the *G. soja* allele had a positive effect (23 cm), with the remaining QTL all displaying negative effects for the *G. soja* allele (Tables 2.5 and 2.6). This combination of positive and negative effects for *G. soja* alleles could explain the transgressive segregation of some RILs both above and below either parent. The QTL for MSL detected in this

study correlated with QTL found in other studies. Using their chromosome segment substitution line population, two of the QTL found by Wang et al. (2013) were also on chr 4 and 12, albeit at different locations on the chromosomes. *qMSL-4, qMSL-13,* and *qMSL-17* were also detected in *G. max* x *G. soja* populations in Kuroda et al. (2013). As part of their study, Zhou et al. (2015) performed a GWAS analysis to detect associations between copy number variation (CNV) and plant height. A CNV located at 26.6 Mb on chr 12 displayed a significant association with plant height and also overlapped with a CNV selection signal between *G. max* and *G. soja*. This CNV could possibly be related to *qMSL-12* in PI479.

Growth Habit

The two RIL populations only shared one QTL (*qGH-6*) associated with growth habit (Table 2.4). Like other growth traits, growth habit was correlated with maturity, and *E1* explained a large percentage of the variation (16%-25%). Two additional QTL were identified in PI468 (*qGH-1* and *qGH-17*) and explained smaller portions of the variation with PVE 17% and 6% respectively. Five QTL were unique to the PI 479752 population, with the largest variation explained by *qGH-6* (16%), followed by *qGH-12* (14%). Some QTL influencing main stem length were also significantly associated with growth habit (*qGH-1* in PI468 and *qGH-12* in PI479). Both QTL corresponded to loci with negative effects on main stem length, and likewise were the only growth habit QTL for which the *G. soja* allele carried a negative effect (towards the *G. max* parent). Other than the *E1* locus, the inability to identify common QTL between the two populations may be due to the importance of genotype specific QTL. However, two QTL (*qGH-2*, and *qGH-18*) were also detected in a *G. max* x *G. soja* by measuring a related trait, the number of times the stem wrapped around a support (Liu et al., 2007).

Lodging

Maturity also played a role in the degree of lodging as *qLD-6* correlates with the location of *E1* but with importance dependent on the population. In PI468, *qLD-6* was the only QTL detected with a PVE of 18%, while in PI479 it played a smaller role with a PVE of 5%. Interestingly, the *G. soja* allele at *qLD-6* increased lodging, while the *G. soja* allele at the same region on chr 6 decreased main stem length and growth habit score. The most important QTL in PI479 were *qLD-2* and *qLD-19-1* (both PVE 10%), neither of which were significant in PI468. Four additional QTL were only detected when the analysis was performed on the combined

populations, two of which occurred on chromosomes containing a lodging QTL in PI479 (*qLD-13-2* and *qLD-19-2*).

100-Seed Weight

Seed weight is one of the most distinctive traits differentiating the wild and domesticated soybean. While some other important domestication-related traits have been found to be controlled by one or a few major QTL (shattering and hard-seededness), 100-seed weight appears to be controlled by the cumulative effects of many QTL with relatively small effects. In spite of the polygenic nature of the trait, a large proportion of the phenotypic variance (85.7%) could still be attributed to genetic components (Table 2.2). One QTL, *qSW-17-1*, was detected within both populations and displayed the largest PVE (16%-19%). However, the additive effect of a *G. soja* allele at this locus was still only a modest -0.3 g to -0.5 g, a small change given the 15 g difference between the *G. max* and *G. soja* parents. Mapping 100-seed weight QTL in the smaller PI 468916 population was ineffective as only two QTL were identified (*qSW-17-1* and *qSW-1-1*) and only 30% of the variance was explained by the full QTL model (Table 2.3). Mapping within PI479 detected more QTL (10) and explained a larger portion of the variance (58%). Combining both populations produced the QTL model that explained the largest portion of variance (68%) and yielded the most loci (14), with five of the QTL only detected in the combined populations. Two separate QTL were detected on chr 1 between PI468 (*qSW-1-1*) and the combined populations (*qSW-1-2*).

Due to its economic importance and significance for domestication, numerous studies have been conducted to identify QTL affecting 100-seed weight. The SoyBase database has records of over 200 associations with 100-seed weight reported in the literature, which were present on all 20 chromosomes (Grant et al., 2010). Among these, a QTL in the vicinity of *qSW-17-1* is frequently detected between *G. max* and *G. soja* (Liu et al., 2007; Zhou et al., 2015; Kulkarni et al., 2016). Many of the other QTL from this present study were also detected in *G. max* x *G. soja* populations, including *qSW-1-2* and *qSW-12* (Yan et al., 2014), *qSW-15* and *qSW-19* (Kulkarni et al., 2016), and *qSW18* (Xin et al., 2016). Although there is agreement between the QTLs detected in this study with others, no two populations reveal the same set of QTL for seed weight. Even *qSW-17-1*, which was the only locus to be detected in both PI468 and PI479, was not detected in all *G. max* x *G. soja* populations (Yan et al., 2014; Han et al., 2016) or was only identified in certain environments (Xin et al., 2016; Kulkarni et al., 2016). The lack of QTL

that are universally detected or have large effects would appear to indicate that the increase in seed size through domestication was not caused by selection for a single gene, but instead was due to the accumulation of many QTL.

Leaflet Size and Shape

While the QTL controlling leaflet size identified in PI468 were also detected in PI479, the PVE of the loci varied by population (Table 2.4). *qLSZ-1* explained the largest portion of variance (26%), followed by *qLSZ-6* (13%) and *qLSZ-3* (10%). In PI479 however, *qLSZ-12* had the largest PVE (21%) and the remaining 10 QTL individually explained under 10% PVE. The majority of leaflet size QTL have previously been identified in other studies (Keim et al., 1990a; Orf et al., 1999; Kim et al., 2005; Chen et al., 2007). The loci on chr 3, 4, 6, and 12 did not correlate with any QTL listed in SoyBase and may represent novel QTL for this trait.

Interestingly, the strongest correlation between any two traits occurred between leaflet size and stem diameter. Presumably, these traits would be unrelated, but both are associated with the *G. soja* accessions used in these populations (small stem diameters and small leaflets). Many leaflet size QTL map to the same position as QTL for main stem length or stem diameter, including those on chr 4, 6, 7, 12, 13, 17, 18, and 19. *qLSZ-3* was not listed in SoyBase and was not associated with any other traits measured in this study, which indicates that this is potentially a novel QTL for leaflet size.

Leaflet shape was controlled by fewer QTL than leaflet size. Only one loci, *qLSH-6,* was significant in the PI 468916 population, and only slightly above the significance threshold. The lack of QTL in PI468 is not surprising given the limited range in distribution of the RILs and the smaller difference between the parents of the population compared to PI479 (Table 2.1 and Fig 2.1). Four QTL were detected in PI479 which cumulatively explained 53% of the variation (Table 2.3). *qLSH-16* and *qLSH-6-1* carried the largest PVE with 23% and 17% respectively, and a significant interaction between them (Tables 2.4 and 2.7). Both of these QTL have previously been identified in *G. max* (Kim et al., 2005; Jun et al., 2014). Among the leaflet shape QTL in this study, *qLSH-17* has not been previously described and is located in a region associated with six different traits, including leaflet size. A qualitative gene controlling leaflet shape, *Ln*, was identified in soybean and was determined to be located on chr 20 (Jeong et al., 2011). Although this is a major gene for the narrow leaflet trait, additional loci also contribute to a quantitative inheritance of the trait. No QTL were detected on chr 20 in this study and no QTL reported in

SoyBase correlate with the *Ln* locus, which indicates that the variation measured in this study and others was not due to this gene and there are other genes affecting narrow leaflets in *G. soja* that may not be in *G. max*.

Pubescence Form

The pubescence of soybeans exhibit variation for several different traits including color, density, presence/absence, and orientation on the upper-side of the leaf (pubescence form). In *G. max*, three categories have historically been used to characterize the latter trait—erect, semi-appressed, and appressed. However, a fourth category (very appressed) is also used by the USDA Soybean Germplasm Collection to differentiate genotypes with extreme levels of appressed pubescence in some *G. soja* accessions. While appressed pubescence is characterized by trichomes mostly flat on the leaf surface, 'very appressed' phenotypes have trichomes completely pressed to the surface and often appear glabrous without close inspection. The three traditional categories have been determined to be controlled by two epistatic loci (Bernard, 1975). Erect pubescence is conditioned by *Pa1*, semi-appressed by *pa1 Pa2*, and appressed by *pa1 pa2* (Palmer et al., 2004).

Using a categorical scoring system to enable QTL mapping, three QTL were detected in PI468 and two in PI479 with high proportions of variation explained by the models (72% and 81% respectively). Both *G. soja* parents displayed the very-appressed phenotype, but there were differences in the QTL detected. While *qPB-12* and *qPB-13* were detected in both populations, *qPB-1* was only detected in PI468. *qPB-13* correlates with the location of *Pa2* on chr 13, which is the locus that controls the change from semi-appressed to appressed (Lee et al., 1999). However, the locus that controls the erect pubescence phenotype (*Pa1*) was mapped by Lee et al. (1999) to chr 11 and was not detected above the significance threshold in these populations. Instead, a possible alternative location for *Pa1* may be on chr 12 (*qPB-12*).

To study the interaction of *qPB-12* and *qPB-13*, the allele calls for each trait category were tabulated in Table 2.8. The erect phenotype was clearly conditioned by the combination of *G. max* alleles at both QTL (*qPB-12m* and *qPB-13m*), while the very appressed phenotype is caused by *G. soja* alleles at each locus (*qPB-12s* and *qPB-13s*). The semi-appressed and appressed categories are less clear, but appear to be due to combinations of the *G. max* and *G. soja* alleles at the two loci. In PI479, lines characterized as semi-appressed tended to carry *qPB-12m* with *qPB-13s*, and appressed phenotypes were due to *qPB-12s* with *qPB-13m* (Table 2.8).

However, in PI468, the semi-appressed phenotypes were split between the two combinations of alleles, while the appressed phenotypes were equally represented by *qPB-12s qPB-13m* and *qPB-12s qPB-13s*.

Assuming that *qPB-12* and *qPB-13* are in fact *Pa1* and *Pa2* respectively, these results would indicate that it is possible to phenotypically distinguish *Pa1 Pa2* from *Pa1 pa2* when the *pa2* allele is derived from *G. soja* accessions with very appressed pubescence. It is possible that these two *G. soja* accessions carry an alternative allele to *pa2* which confers the very appressed pubescence form and for which *Pa1* is only partially dominant. This is also reflected by the small interaction effect between the QTL, which is characteristic of additive loci. Crosses between genotypes with appressed and very appressed pubescence could be developed to confirm the alternate *pa2* allele.

Overlapping QTL

QTL for the 11 different traits were located on all 20 chromosomes but with unequal representation (Table 2.9, Fig. 2.4-2.6). Chromosome 6 contained the most QTL (11), while chr 3, chr 5, and chr 10 contained the fewest with only one each. However, the abundance of loci was often due to the association of a specific locus with multiple traits. One example of this can be observed on chr 6, for which a QTL in the vicinity of the *E1* locus was associated with nine different traits (excluding leaflet size and pubescence form). Based on overlapping or proximal 1.5-LOD support intervals, 36 genomic regions containing QTL were detected, with 1 to 3 regions on each chromosome (Table 2.9). Some regions contained QTL for multiple traits within a narrow interval (e.g. region 18 on chr 12 between 36.90-37.53 Mb), while other regions spanned a large interval (e.g. region 20 on chr 13 between 30.50-39.25 Mb). The narrow intervals are most likely the result of a single QTL with pleiotropic effects, while the broad genomic regions may potentially contain multiple QTL that are more difficult to separate with the limited precision caused by the extended LD in soybean. The QTL in region 9 on chr 6 are likely caused by the *E1* maturity gene. Although the 1.5-LOD intervals for the QTL in this region encompass a large segment of the chromosome (21.26 Mb), this is likely due to the reduced recombination in the region as opposed to multiple underlying genes (Fig. 2.3).

Region 34 on chr 19 contained QTL from seven different traits. With the exception of *qR8-19*, all had nearly identical 1.5-LOD support intervals and positions (Table 2.9). This provides evidences that these QTL likely originate from a single locus with pleiotropic effects as

opposed to multiple linked loci. Most of these QTL were only detected in the PI 479752 population (*qLSZ-19* was only mapped in the combined set of RILs) and the effects of the *G. soja* alleles were in the direction of the *G. soja* parent (Table 2.6). Some traits represented within region 34 would be expected to be phenotypically related, such as stem diameter, growth habit, and lodging. Plants with small stem diameters and vining plants are more prone to lodging. Interestingly, a phenotypic relationship between the other traits in this region is not as obvious. Although leaflet size and seed weight could also be affected by a gene that generally affects the size of plant parts, explaining leaflet shape and time of maturity as part of this group of traits is more difficult.

Considering the six traits directly related to the domestication of soybean (stem diameter, main stem length, shattering, growth habit, lodging, and seed weight), there were a total of 59 QTL in 32 genomic regions on all chromosomes except chr 3. This differs from the findings by Liu et al. (2007) who detected domestication-related QTL on only six chromosomes. However, this difference is plausibly a result of the larger population sizes used in the current study. In the PI 468916 population, which was similar in size to the Liu et al. (2007) study, QTL were only detected on seven different chromosomes. The large number of QTL for domestication-related traits should be expected given the abundance of selective sweeps detected by several different studies (Lam et al., 2010; Li et al., 2013; Zhou et al., 2015). While the domestication selective sweeps have been detected on all 20 chromosomes, they are not randomly distributed (Li et al., 2013). This is also consistent with the findings from the present QTL mapping study, in which 12 out of 32 genomic regions harbor QTL for multiple domestication-related traits. Of the 14 QTL that were only associated with a single trait, four were associated with 100-seed weight, and four were associated with lodging (Table 2.9). These QTL only explained a small portion of the variation for their respective traits, with *qSW-12* displaying the largest PVE at 11%.

Among the six DRT measured in this study, four were related to plant growth traits (stem diameter, main stem length, growth habit, and lodging). Although QTL for these traits were frequently detected together in the same region, only region 9 on chr 6 was associated with all four traits (Table 2.9). Region 9 contains the *E1* locus for maturity and flowering time, which also affects growth traits in indeterminate soybean. Later maturing lines tended to exhibit larger stem diameters, main stem lengths, and growth habit ratings, with a smaller but positive correlation with lodging as well (Fig. 2.2). The *E1* maturity gene is likely the underlying gene for

this region, although additional closely linked QTL should not be ruled out. Only ten QTL showed no overlap among the four growth traits, *qMSL-4, qMSL-9, qSD-7, qSD-14*, *qLD-10*, *qLD-11*, *qLD-12, qLD-13-1, qLD-13-2,* and *qLD-19-2*. This demonstrates the correlated nature between the traits. Of the ten QTL, *qMSL-9, qSD-14*, *qLD-10*, *qLD-13-1*, *qLD-13-2,* and *qLD-19-2* were detected in regions without QTL for any other traits, which indicates that these four QTL only affect one trait.

CONCLUSIONS

Several studies have been conducted to map qualitative traits using *G. max* x *G. soja* populations (Keim et al., 1990a; Liu et al., 2007; Yan et al., 2014). However, the studies published to date have been limited by marker density, population size, or both. In the current study, two *G. max* x *G. soja* populations with 151 and 510 RILs were genotyped through GBS to obtain approximately 17,000 polymorphic SNP markers in each population. The combination of a larger mapping population and a denser set of markers improved the ability to detect additional QTL. In the PI 468916 population with 151 RILs, significant QTL were limited to those with large PVE and no more than three QTL were detected for any trait. This scenario is similar to what was found in the Liu et al. (2007) study and others that were constrained by population size. Using the larger PI 479752 population, up to eleven loci were detected, including QTL only explaining a small portion of the phenotypic variance.

Among the traits measured, growth habit and pubescence form utilized novel phenotyping methodologies to describe the variation. Previously, vining tendencies have typically been quantitatively measured by dissecting the trait into twining habit and maximum internode length. Applying a visual scoring system towards the overall growth habit of the plant allows the incorporation of multiple traits that would otherwise be difficult to measure on a large scale. Pubescence form has also traditionally been evaluated on a scale with three classes, while the USDA Soybean Germplasm Collection also documents a fourth class (very appressed). This study was the first to incorporate all four classes of pubescence form. Based on these findings, the very appressed pubescence appears to be controlled by the same two *Pa* genes, *pa1 pa2*. Potentially, an alternative allele at *pa2* may cause very appressed pubescence in association with *pa1*. The ability to phenotypically differentiate *Pa1 Pa2* from *Pa1 pa2* would also imply that this putative *pa2* allele interacts in a more additive nature than the previously described allele.

Three of the six domestication-related traits (main stem length, shattering, and growth habit) were controlled by one or two QTL that explain over 20% of the variation with several minor QTL. The remaining domestication traits lacked major QTL and were instead controlled by the minor QTL alone. The two *G. max* x *G. soja* populations had limited overlap in QTL for domestication-related traits, with only *qSD-1*, *qSD-13, qLD-6, qMSL-6, qGH-6, qSH-16,* and *qSW-17-1* occurring in both. However, considerably fewer DRT QTL were detected in PI468 overall (12 compared to 43 in PI479) due to reduced statistical power with fewer RILs. The majority of genomic regions with domestication QTL were associated with only one trait, and only 12 out of 32 regions were associated with multiple domestication-related traits. This is a similar result as observed by Liu et al. (2007) who found that only four regions contained QTL for multiple traits. The QTL identified in this study represent important regions in the domestication of soybean from *G. soja*. QTL that were detected in both *G. max* x *G. soja* populations may be examples of key loci that were selected early in the domestication process.

TABLES

Table 2.1. Trait means and ranges for the parents and RIL progeny. Standard deviations are given in parentheses.

Entry	R1[†] (days)	R1DAP[‡] (days)	R8[§] (days)	Stem Diameter (mm)	Main Stem Length (cm)	Shattering (rating)	Growth Habit (rating)	Lodging (rating)
PI 468916	65.5 (3.8)	69.2 (3.3)	114.8 (1)	2.68 (0.66)	73.5 (26.8)	6 (0.2)	4.8 (0.3)	9 (0.4)
PI 479752	41.8 (9.3)	45.4 (7.9)	101.6 (2.5)	2.32 (0.56)	81.2 (34.3)	5.6 (1)	4.6 (1)	9 (0)
Williams 82	40 (6.7)	43.7 (6.3)	128.2 (5.2)	8.68 (0.79)	97 (18.8)	1.1 (0.3)	1 (0)	2.8 (0.9)
RIL								
Mean	43.5 (12.4)	47.2 (12.4)	116 (12.3)	4.59 (1.09)	103.7 (38.7)	3.6 (1.3)	3 (0.9)	7.6 (0.9)
Range	25.5-76.6	29.2-80.3	87.1-147.1	2.37-8.65	24.5-200.8	1-6	0.9-5	2.9-8.9

Entry	100-Seed Weight (g)	Leaf Size (rating)	Leaf Shape (rating)
PI 468916	1.3 (0.3)	1.2 (0.3)	1.9 (0)
PI 479752	1.3 (0.3)	1.3 (0.3)	2.8 (0.3)
Williams 82	16.5 (1.9)	4.9 (0.4)	1 (0)
RIL			
Mean	4.3 (1.4)	3 (0.7)	1.5 (0.5)
Range	2-11.2	1.3-4.8	1-3

[†]Days to flowering, as days after May 31

[‡]Days to flowering, as days after planting

[§]Days to maturity, as days after May 31

Table 2.2. Variance components of phenotypic traits. RIL (genotype) considered fixed while all other components are random.

	R1[†] (days)	R1DAP[‡] (days)	R8[§] (days)	Stem Diameter (mm)	Main Stem Length (cm)	Shattering (rating)	Growth Habit (rating)	Lodging (rating)	100-Seed Weight (g)	Leaf Size (rating)	Leaf Shape (rating)
Env	22.5	3.6	1	8.4	9.1	18.4	2	2.4	3.1	12	4.3
Block	0.1	0.1	0.8	0.2	2.8	0.1	0.1	0.4	0.3	4.6	0.3
RIL	71.7	92.2	90.3	67.9	63.6	50.1	73	52.4	85.7	55.7	48
RIL x Env	3.4	1.9	3.2	3.9	6.2	14.6	5.4	13	3.6	6	17.1
Error	2.3	2.2	4.7	19.7	18.3	16.7	19.5	31.8	7.3	21.7	30.3

[†]Days to flowering, as days after May 31

[‡]Days to flowering, as days after planting

[§]Days to maturity, as days after May 31

Table 2.3. Summaries of the QTL models, with the number of QTL and the phenotypic variance explained by the model (R^2). QTL mapping was performed within each population (PI468 and PI479) and across both populations (combined) while using the pedigree as a fixed effect. The number of QTL unique to a specific population, or only detected in the combined model is given in parentheses.

	PI468		PI479		Combined	
Trait	# QTL	R^2	# QTL	R^2	# QTL	R^2
Flowering Date (R1DAP)	1	0.79	5 (4)	0.86	6 (1)	0.81
Maturity Date (R8)	1	0.54	9 (8)	0.79	9	0.77
Stem Diameter	2	0.28	10 (8)	0.62	10	0.6
Main Stem Length	3 (2)	0.60	4 (3)	0.64	8 (2)	0.66
Shattering	1	0.23	6 (5)	0.75	6	0.7
Growth Habit	3 (2)	0.48	6 (5)	0.52	7 (1)	0.49
Lodging	1	0.18	7 (6)	0.41	10 (4)	0.45
100-Seed Weight	2 (1)	0.30	10 (9)	0.58	14 (5)	0.68
Leaflet Size	3	0.50	11 (8)	0.64	13 (5)	0.67
Leaflet Shape	1 (1)	0.12	4 (4)	0.53	4	0.45
Pubescence Form	3 (1)	0.72	2	0.81	3	0.75

Table 2.4. Summary of significant QTL and physical position (Mb) in each mapping population. The phenotypic variance explained (PVE) is given for each QTL.

			PI 468916		PI 479752		Combined	
Trait[†]	QTL	Chr[‡]	Mb	PVE	Mb	PVE	Mb	PVE
R1DAP	*qR1DAP-4*	4	-	-	41.44	1	9.76	1
	qR1DAP-6	6	19.51	79	20.02	55	20.26	55
	qR1DAP-7	7	-	-	-	-	3.78	1
	qR1DAP-11	11	-	-	10.61	<1	10.65	1
	qR1DAP-12	12	-	-	5.56	<1	5.38	1
	qR1DAP-16	16	-	-	4.23	22	4.23	12
R8	*qR8-4*	4	-	-	17.68	7	17.68	7
	qR8-6	6	24.72	54	20.47	42	20.27	44
	qR8-7	7	-	-	3.83	5	3.83	4
	qR8-11	11	-	-	10.96	3	10.84	3
	qR8-12	12	-	-	5.38	2	5.47	3
	qR8-13	13	-	-	36.67	1	36.67	2
	qR8-16	16	-	-	4.16	6	4.16	3
	qR8-19	19	-	-	46.22	3	46.22	3
	qR8-20	20	-	-	39.46	3	39.46	2
SD	*qSD-1*	1	48.52	13	48.44	10	48.52	12
	qSD-2	2	-	-	43.56	7	43.56	7
	qSD-6	6	-	-	20.15	11	20.27	12
	qSD-7	7	-	-	4.88	2	4.88	4
	qSD-12	12	-	-	37.22	11	37.22	10
	qSD-13	13	37.08	15	34.02	5	34.03	7
	qSD-14	14	-	-	1.72	3	2.45	5
	qSD-17	17	-	-	12.42	6	12.14	6
	qSD-18	18	-	-	5.27	7	5.13	6
	qSD-19	19	-	-	43.82	16	43.79	15
MSL	*qMSL-1*	1	6.04	25	-	-	6.71	4
	qMSL-4	4	-	-	6.5	2	6.65	3
	qMSL-6	6	38.81	29	20.71	33	20.27	32
	qMSL-9	9	-	-	2.82	3	3.4	4
	qMSL-12	12	-	-	37.18	24	37.18	17
	qMSL-13	13	30.73	14	-	-	31.92	4
	qMSL-17	17	-	-	-	-	10.6	1
SH	*qSH-2*	2	-	-	45.27	4	45.1	3
	qSH-6	6	-	-	19.93	3	20.27	4
	qSH-7	7	-	-	4.28	6	3.83	5
	qSH-15	15	-	-	12.57	6	12.51	4
	qSH-16	16	30.24	23	30.06	53	30.06	39
	qSH-19	19	-	-	37.75	5	37.9	5

Table 2.4 (cont.)

			PI 468916		PI 479752		Combined	
Trait†	QTL	Chr‡	Mb	PVE	Mb	PVE	Mb	PVE
GH	*qGH-1*	1	5.7	17	-	-	-	-
	qGH-2	2	-	-	44.33	9	44.55	8
	qGH-6	6	39.38	25	23.56	16	23.66	17
	qGH-8	8	-	-	-	-	3.2	2
	qGH-12	12	-	-	37.22	14	37.18	10
	qGH-13	13	-	-	39	10	38.94	6
	qGH-17	17	9.08	6	-	-	-	-
	qGH-18	18	-	-	5.76	5	5.81	3
	qGH-19	19	-	-	43.82	3	43.79	4
LD	*qLD-1*	1	-	-	48.51	4	-	-
	qLD-2	2	-	-	43.94	10	43.7	7
	qLD-6	6	20.27	18	21.09	5	31.24	7
	qLD-8	8	-	-	-	-	3.33	4
	qLD-10	10	-	-	-	-	4.76	2
	qLD-11	11	-	-	11.56	4	9.46	2
	qLD-12	12	-	-	5.62	3	5.47	3
	qLD-13-1	13	-	-	42.23	7	41.92	6
	qLD-13-2	13	-	-	-	-	24.09	3
	qLD-19-1	19	-	-	43.82	10	43.79	9
	qLD-19-2	19	-	-	-	-	2.64	4
SW	*qSW-1-1*	1	3.58	12	-	-	-	-
	qSW-1-2	1	-	-	-	-	47.95	4
	qSW-2	2	-	-	41.9	6	41.9	5
	qSW-4-1	4	-	-	2.42	9	2.44	8
	qSW-4-2	4	-	-	40.38	8	40.09	7
	qSW-5	5	-	-	3.88	5	-	-
	qSW-6	6	-	-	-	-	5.41	2
	qSW-9	9	-	-	-	-	35.56	4
	qSW-12	12	-	-	22.93	11	19.04	10
	qSW-14	14	-	-	42.94	5	40.67	5
	qSW-15	15	-	-	-	-	3.73	3
	qSW-17-1	17	7.03	19	9.39	16	8.9	13
	qSW-17-2	17	-	-	14.24	13	14.25	9
	qSW-18	18	-	-	-	-	45.9	5
	qSW-19	19	-	-	43.82	10	43.79	9
	qSW-20	20	-	-	33.52	4	31.37	4

Table 2.4 (cont.)

			PI 468916		PI 479752		Combined	
Trait[†]	QTL	Chr[‡]	Mb	PVE	Mb	PVE	Mb	PVE
LSZ	*qLSZ-1*	1	27.3	26	6.95	5	24.63	8
	qLSZ-3	3	40.57	10	35.09	3	35.65	4
	qLSZ-4-1	4	-	-	-	-	4.83	7
	qLSZ-4-2	4	-	-	7.38	5	9.18	6
	qLSZ-6	6	38.81	13	19.93	9	21.35	10
	qLSZ-7	7	-	-	4.88	3	6.72	4
	qLSZ-9-1	9	-	-	37.67	3	-	-
	qLSZ-9-2	9	-	-	-	-	2.01	4
	qLSZ-12	12	-	-	37.18	21	37.18	15
	qLSZ-13	13	-	-	32.96	7	32.91	7
	qLSZ-14-1	14	-	-	4.82	4	-	-
	qLSZ-14-2	14	-	-	-	-	43.31	1
	qLSZ-17	17	-	-	13.38	5	11.79	4
	qLSZ-18	18	-	-	5.22	5	-	-
	qLSZ-19	19	-	-	-	-	43.79	5
	qLSZ-20	20	-	-	-	-	34.5	2
LSH	*qLSH-6-1*	6	-	-	21.09	17	20.6	12
	qLSH-6-2	6	5.22	12	-	-	-	-
	qLSH-16	16	-	-	27.25	23	27.29	16
	qLSH-17	17	-	-	13.88	8	13.88	7
	qLSH-19	19	-	-	43.82	6	43.79	5
PB	*qPB-1*	1	9.05	43	-	-	25.96	6
	qPB-12	12	37.06	12	37.18	61	37.18	47
	qPB-13	13	32.19	33	31.79	31	31.79	31

[†]R1DAP, first flowering (days after planting); R8, full maturity (days); SD, stem diameter (mm); MSL, main stem length (cm); SH, shattering (rating); GH, growth habit (rating); LD, lodging (rating); SW, 100-seed weight (g); LSZ, leaflet size (rating); LSH, leaflet shape (rating); PB, pubescence form (rating)

[‡]Chromosome

Table 2.5. Significant QTL detected in the Williams 82 x PI 468916 population.

Trait[†]	QTL	Chr[‡]	Position (bp)	1.5-LOD Interval (Mb)	-log(p)	Additive Effect[§]	PVE[¶]
R1DAP	qR1DAP-6	6	19,513,791	18.92 - 20.27	49	13.8	78.7
R8	qR8-6	6	24,716,064	19.22 - 31.24	26.4	10.8	54.1
SD	qSD-13	13	37,082,064	34.93 - 37.38	7.2	-0.3	15.3
	qSD-1	1	48,523,860	41.65 - 49.14	6	-0.3	12.9
MSL	qMSL-6	6	38,814,963	19.51 - 39.38	18.5	23.3	28.9
	qMSL-1	1	6,041,028	5.02 - 23.72	16.6	-19.8	24.9
	qMSL-13	13	30,726,049	30.45 - 31.79	5.5	-11.2	14.4
SH	qSH-16	16	30,236,035	29.62 - 31.24	9.4	0.6	22.7
GH	qGH-6	6	39,379,331	19.22 - 41.44	13.7	0.5	24.9
	qGH-1	1	5,696,885	5.02 - 7.60	9.2	-0.5	16.9
	qGH-17	17	9,078,271	8.19 - 9.65	5.5	0.3	5.9
LD	qLD-6	6	20,272,282	18.33 - 39.38	6.3	0.4	18.5
SW	qSW-17-1	17	7,031,145	6.08 - 7.65	8.8	-0.5	19.4
	qSW-1-1	1	3,582,277	1.80 - 4.57	5.2	-0.4	12
LSZ	qLSZ-1	1	27,303,439	5.17 - 35.65	14.8	-0.3	26
	qLSZ-6	6	38,814,963	19.51 - 41.07	8.2	0.3	13
	qLSZ-3	3	40,573,425	38.96 - 41.85	7	-0.2	10.4
LSH	qLSH-6-2	6	5,218,411	5.06 - 6.12	5.1	0.1	12
PB	qPB-1	1	9,047,275	5.02 - 11.45	23.8	0.6	43
	qPB-13	13	32,191,744	31.79 - 32.36	13.9	0.5	32.7
	qPB-12	12	37,055,412	36.76 - 37.77	7.4	0.3	11.9

[†]R1DAP, first flowering (days after planting); R8, full maturity (days); SD, stem diameter (mm); MSL, main stem length (cm); SH, shattering (rating); GH, growth habit (rating); LD, lodging (rating); SW, 100-seed weight (g); LSZ, leaflet size (rating); LSH, leaflet shape (rating); PB, pubescence form (rating)

[‡]Chromosome

[§]Additive effect of a *G. soja* allele at the locus. See trait for units.

[¶]Phenotypic variance explained

Table 2.6. Significant QTL detected in the Williams 82 x PI 479752 population.

Trait[†]	QTL	Chr[‡]	Position (bp)	1.5-LOD Interval (Mb)	-log(p)	Additive Effect[§]	PVE[¶]
R1DAP	*qR1DAP-6*	6	20,016,081	19.90 - 20.84	172.5	9.5	54.8
	qR1DAP-16	16	4,232,473	4.16 - 4.33	95.3	-3.2	22.2
	qR1DAP-4	4	41,437,276	9.76 - 43.77	14.7	-3.5	1.3
	qR1DAP-12	12	5,561,658	5.38 - 5.69	21.8	2.1	0.2
	qR1DAP-11	11	10,611,634	10.30 - 10.91	11.1	0.2	0.1
R8	*qR8-6*	6	20,469,817	20.02 - 21.09	114.9	7.4	42.9
	qR8-4	4	17,684,571	14.89 - 42.56	37.1	-4.8	7.2
	qR8-16	16	4,159,052	3.26 - 4.25	21.2	-2.5	6.2
	qR8-7	7	3,829,809	3.53 - 4.41	23.5	-2.7	4.9
	qR8-19	19	46,224,374	44.96 - 48.00	6.6	-1.5	3.3
	qR8-11	11	10,958,511	10.72 - 11.05	23.5	2.8	2.8
	qR8-20	20	39,455,640	38.82 - 40.72	10.1	-1.8	2.6
	qR8-12	12	5,379,500	5.32 - 5.69	21.3	3	1.8
	qR8-13	13	36,668,674	36.19 - 39.69	10.2	-1.6	1.4
SD	*qSD-19*	19	43,823,792	43.79 - 44.25	36.6	-0.3	15.7
	qSD-12	12	37,224,910	37.06 - 37.41	25.1	-0.3	11.5
	qSD-6	6	20,149,620	19.40 - 22.26	25.1	0.3	11
	qSD-1	1	48,443,918	47.74 - 49.07	13.2	-0.2	9.9
	qSD-18	18	5,272,063	5.07 - 5.56	16.8	-0.2	6.9
	qSD-2	2	43,562,617	43.40 - 43.94	11.2	-0.2	6.7
	qSD-17	17	12,421,516	12.08 - 12.92	10.4	-0.2	6.1
	qSD-13	13	34,023,572	32.91 - 34.03	7.5	-0.2	5.2
	qSD-14	14	1,722,233	1.24 - 2.45	9.2	-0.2	2.9
	qSD-7	7	4,875,404	3.95 - 5.45	8.7	-0.2	2.3
MSL	*qMSL-6*	6	20,714,818	19.93 - 21.27	71.2	23.4	33.4
	qMSL-12	12	37,178,672	37.06 - 37.41	57.7	-19.1	24.1
	qMSL-9	9	2,823,269	0.45 - 3.50	7.1	-5.8	3.4
	qMSL-4	4	6,495,299	6.46 - 6.65	11.8	-7.4	2.3
SH	*qSH-16*	16	30,059,450	29.90 - 30.19	120.7	0.8	52.6
	qSH-7	7	4,280,914	3.78 - 4.61	21.8	0.3	5.7
	qSH-15	15	12,565,395	12.23 - 12.60	16.8	0.1	5.6
	qSH-19	19	37,745,363	37.61 - 38.23	21.9	0.3	5.1
	qSH-2	2	45,273,334	44.81 - 45.44	9.6	0.2	4.2
	qSH-6	6	19,925,553	19.10 - 21.70	16	-0.2	3.3
GH	*qGH-6*	6	23,558,078	19.88 - 36.33	30.6	0.4	16.4
	qGH-12	12	37,224,910	36.93 - 37.41	29.4	-0.4	13.8
	qGH-13	13	39,002,190	38.45 - 39.19	16.3	0.2	10
	qGH-2	2	44,331,133	43.56 - 44.77	18.8	0.2	8.6
	qGH-18	18	5,755,011	5.23 - 5.93	10.2	0.2	5.3
	qGH-19	19	43,823,792	42.26 - 45.93	7.1	0.2	3.2

Table 2.6. (cont.)

Trait[†]	QTL	Chr[‡]	Position (bp)	1.5-LOD Interval (Mb)	-log(p)	Additive Effect[§]	PVE[⁋]
LD	*qLD-19-1*	19	43,823,792	43.64 - 44.37	18.2	0.2	10.7
	qLD-2	2	43,941,599	43.56 - 44.77	14	0.2	9.9
	qLD-13-1	13	42,226,516	41.63 - 43.15	12.9	0.2	7
	qLD-6	6	21,092,826	16.99 - 39.06	9.4	0.2	4.7
	qLD-11	11	11,560,622	11.19 - 13.86	7.8	0.2	4.6
	qLD-1	1	48,505,842	47.74 - 49.06	7.2	0.2	4.3
	qLD-12	12	5,617,171	5.32 - 6.38	7	0.2	3.1
SW	*qSW-17-1*	17	9,394,030	8.81 - 10.17	35.2	-0.3	15.8
	qSW-17-2	17	14,240,110	14.03 - 15.28	9.1	-0.3	13.2
	qSW-12	12	22,927,593	11.92 - 32.27	19.3	-0.3	11.3
	qSW-19	19	43,823,792	43.26 - 44.25	27.5	-0.3	10.3
	qSW-4-1	4	2,415,968	1.80 - 2.82	16.4	-0.2	9.5
	qSW-4-2	4	40,375,588	12.63 - 43.35	14	-0.3	8.4
	qSW-2	2	41,897,975	41.54 - 42.73	7.6	-0.2	6.1
	qSW-14	14	42,936,962	35.41 - 44.36	10.4	-0.2	5.2
	qSW-5	5	3,875,852	3.52 - 4.28	10.5	-0.2	4.8
	qSW-20	20	33,524,399	27.90 - 33.74	8.5	-0.2	3.7
LSZ	*qLSZ-12*	12	37,178,672	37.06 - 37.41	50.9	-0.3	21.8
	qLSZ-6	6	19,925,553	19.40 - 21.70	24.9	0.2	9.4
	qLSZ-13	13	32,959,251	31.79 - 34.35	10.8	-0.1	6.7
	qLSZ-17	17	13,380,584	11.79 - 15.28	13.9	-0.1	5.3
	qLSZ-4-2	4	7,383,917	4.50 - 9.23	12.6	-0.1	5.2
	qLSZ-18	18	5,215,223	5.01 - 5.32	17.9	-0.1	5.1
	qLSZ-1	1	6,946,727	3.66 - 27.15	7.9	-0.1	4.9
	qLSZ-14	14	4,822,711	4.52 - 5.75	7.2	-0.1	3.8
	qLSZ-9-1	9	37,667,446	29.36 - 40.28	6.5	-0.1	3.4
	qLSZ-7	7	4,875,404	4.61 - 5.90	10.4	-0.1	3.3
	qLSZ-3	3	35,090,391	34.69 - 37.06	8.3	-0.1	2.7
LSH	*qLSH-16*	16	27,245,155	26.92 - 27.31	45.1	0.4	22.7
	qLSH-6-1	6	21,092,826	19.93 - 21.70	33.2	-0.2	17.4
	qLSH-17	17	13,878,376	13.52 - 14.17	18.8	0.2	7.7
	qLSH-19	19	43,823,792	43.39 - 44.60	11.5	0.1	6.2
PB	*qPB-12*	12	37,178,672	37.06 - 37.41	133.4	0.7	60.9
	qPB-13	13	31,792,004	31.61 - 31.92	67.5	0.4	31.5

[†]R1DAP, first flowering (days after planting); R8, full maturity (days); SD, stem diameter (mm); MSL, main stem length (cm); SH, shattering (rating); GH, growth habit (rating); LD, lodging (rating); SW, 100-seed weight (g); LSZ, leaflet size (rating); LSH, leaflet shape (rating); PB, pubescence form (rating)

[‡]Chromosome

[§]Additive effect of a *G. soja* allele at the locus. See trait for units.

[⁋]Phenotypic variance explained

Table 2.7. Epistatic interactions among QTL in Williams 82 x PI 479752.

Trait[†]	QTLs		-log10(p)	PVE[‡]
R1DAP	qR1DAP-6	qR1DAP-16	27.8	4.8
	qR1DAP-6	qR1DAP-12	3.1	0.7
	qR1DAP-6	qR1DAP-4	4.3	<0.1
	qR1DAP-6	qR1DAP-11	7.1	1.8
	qR1DAP-16	qR1DAP-12	3.2	0.7
	qR1DAP-16	qR1DAP-4	5.8	0.9
R8	qR8-6	qR8-4	5.1	0.5
SH	qSH-16	qSH-15	4.6	0.7
LSH	qLSH-16	qLSH-6	3.2	1.5
PB	pPB-12	pPB-13	4.7	0.8

[†]R1DAP, first flowering (days after planting); R8, full maturity (days); SH, shattering (rating); LSH, leaflet shape (rating); PB, pubescence form (rating)

[‡]Phenotypic variance explained

Table 2.8. Counts of pubescence form for each genotype at *qPB-12* and *qPB-13*

	Alleles[†]	Pubescence Classes[‡]			
	qPB-12 / qPB-13	E	SA	A	VA
PI479[§]	*m / m*	142	17	0	1
	m / s	4	80	2	2
	s / m	1	30	29	7
	s / s	0	1	3	69
PI468[‖]	*m / m*	24	7	0	0
	m / s	3	18	1	6
	s / m	8	12	4	1
	s / s	0	4	4	10

[†]*m*, homozygous for *G. max* allele; *s*, homozygous for *G. soja* allele

[‡]E, erect; SA, semi-appressed; A, appressed; VA, very appressed

[§]Williams 82 x PI 479752

[‖]Williams 82 x PI 468916

Table 2.9. Genomic regions containing QTL. The physical positions and 1.5-LOD support intervals are reported as the averages across the three mapping populations (PI468, PI479, and the combined RILs).

Region	Trait[†]	QTL	Chr[‡]	Position (Mb)	1.5-LOD Interval (Mb)
1	SW	*qSW-1-1*	1	3.58	1.80 - 4.57
	GH	*qGH-1*	1	5.70	5.02 - 7.60
	MSL	*qMSL-1*	1	6.38	5.09 - 25.01
	PB	*qPB-1*	1	17.50	5.53 - 23.97
	LSZ	*qLSZ-1*	1	19.63	6.23 - 29.81
2	SW	*qSW-1-2*	1	47.95	46.81 - 48.52
	SD	*qSD-1*	1	48.50	45.73 - 49.00
	LD	*qLD-1*	1	48.51	47.74 - 49.06
3	SW	*qSW-2*	2	41.90	41.50 - 42.47
	SD	*qSD-2*	2	43.56	43.40 - 43.94
	LD	*qLD-2*	2	43.82	43.53 - 44.77
	GH	*qGH-2*	2	44.44	43.63 - 44.79
	SH	*qSH-2*	2	45.19	44.68 - 45.46
4	LSZ	*qLSZ-3*	3	37.10	36.17 - 38.87
5	SW	*qSW-4-1*	4	2.43	1.80 - 2.95
	LSZ	*qLSZ-4-1*	4	4.83	4.28 - 4.97
	MSL	*qMSL-4*	4	6.57	6.46 - 6.79
	LSZ	*qLSZ-4-2*	4	8.28	6.83 - 9.23
6	R8	*qR8-4*	4	17.68	14.80 - 30.73
	R1DAP	*qR1DAP-4*	4	25.60	9.69 - 29.69
	SW	*qSW-4-2*	4	40.23	12.66 - 43.35
7	SW	*qSW-5*	5	3.88	3.52 - 4.28
8	LSH	*qLSH-6-2*	6	5.22	5.06 - 6.12
	SW	*qSW-6*	6	5.41	4.59 - 7.63
9	R1DAP	*qR1DAP-6*	6	19.93	19.35 - 20.57
	SH	*qSH-6*	6	20.10	19.16 - 21.94
	SD	*qSD-6*	6	20.21	19.83 - 25.72
	LSH	*qLSH-6-1*	6	20.85	20.09 - 21.69
	R8	*qR8-6*	6	21.82	19.83 - 24.31
	LD	*qLD-6*	6	24.20	18.01 - 39.27
	MSL	*qMSL-6*	6	26.60	19.55 - 27.61
	LSZ	*qLSZ-6*	6	26.70	19.38 - 28.32
	GH	*qGH-6*	6	28.87	19.44 - 36.34
10	R1DAP	*qR1DAP-7*	7	3.78	3.45 - 4.00
	R8	*qR8-7*	7	3.83	3.53 - 4.20
	SH	*qSH-7*	7	4.06	3.62 - 4.61
	SD	*qSD-7*	7	4.88	3.97 - 5.45
	LSZ	*qLSZ-7*	7	5.80	4.68 - 6.57

Table 2.9. (cont.)

Region	Trait[†]	QTL	Chr[‡]	Position (Mb)	1.5-LOD Interval (Mb)
11	GH	*qGH-8*	8	3.20	2.30 - 3.70
	LD	*qLD-8*	8	3.33	2.50 - 4.15
12	LSZ	*qLSZ-9-2*	9	2.01	1.20 - 3.48
	MSL	*qMSL-9*	9	3.11	1.64 - 3.49
13	SW	*qSW-9*	9	35.56	22.08 - 36.59
	LSZ	*qLSZ-9-1*	9	37.67	29.36 - 40.28
14	LD	*qLD-10*	10	4.76	4.55 - 5.97
15	LD	*qLD-11*	11	10.51	9.93 - 12.26
	R1DAP	*qR1DAP-11*	11	10.63	10.46 - 10.98
	R8	*qR8-11*	11	10.90	10.69 - 11.05
16	R8	*qR8-12*	12	5.42	5.32 - 6.04
	R1DAP	*qR1DAP-12*	12	5.47	5.35 - 5.69
	LD	*qLD-12*	12	5.54	5.32 - 6.38
17	SW	*qSW-12*	12	20.99	15.22 - 31.78
18	PB	*qPB-12*	12	37.14	36.90 - 37.53
	LSZ	*qLSZ-12*	12	37.18	36.97 - 37.32
	MSL	*qMSL-12*	12	37.18	36.97 - 37.41
	GH	*qGH-12*	12	37.20	36.87 - 37.41
	SD	*qSD-12*	12	37.22	37.12 - 37.41
19	LD	*qLD-13-2*	13	24.09	23.26 - 27.08
20	MSL	*qMSL-13*	13	31.32	30.50 - 32.21
	PB	*qPB-13*	13	31.93	31.67 - 32.22
	LSZ	*qLSZ-13*	13	32.94	31.70 - 34.35
	SD	*qSD-13*	13	35.05	33.91 - 35.48
	R8	*qR8-13*	13	36.67	36.24 - 38.84
	GH	*qGH-13*	13	38.97	38.45 - 39.25
21	LD	*qLD-13-1*	13	42.07	41.62 - 43.07
22	SD	*qSD-14*	14	2.09	1.48 - 2.79
23	LSZ	*qLSZ-14-1*	14	4.82	4.52 - 5.75
24	SW	*qSW-14*	14	41.81	33.39 - 44.46
	LSZ	*qLSZ-14-2*	14	43.31	40.67 - 44.55
25	SW	*qSW-15*	15	3.73	2.74 - 4.57
26	SH	*qSH-15*	15	12.54	12.08 - 12.62
27	R8	*qR8-16*	16	4.16	3.26 - 4.25
	R1DAP	*qR1DAP-16*	16	4.23	4.16 - 4.29
28	LSH	*qLSH-16*	16	27.27	27.04 - 27.31
	SH	*qSH-16*	16	30.12	29.71 - 30.54

Table 2.9. (cont.)

Region	Trait[†]	QTL	Chr[‡]	Position (Mb)	1.5-LOD Interval (Mb)
29	SW	*qSW-17-1*	17	8.44	7.43 - 8.97
	GH	*qGH-17*	17	9.08	8.19 - 9.65
	MSL	*qMSL-17*	17	10.60	8.19 - 11.79
	SD	*qSD-17*	17	12.28	11.94 - 12.74
	LSZ	*qLSZ-17*	17	12.59	10.80 - 14.27
	LSH	*qLSH-17*	17	13.88	13.49 - 14.21
	SW	*qSW-17-2*	17	14.25	13.97 - 15.16
30	SD	*qSD-18*	18	5.20	4.65 - 6.26
	LSZ	*qLSZ-18*	18	5.22	5.01 - 5.32
	GH	*qGH-18*	18	5.78	5.18 - 6.74
31	SW	*qSW-18*	18	45.90	19.93 - 46.74
32	LD	*qLD-19-2*	19	2.64	0.88 - 7.80
33	SH	*qSH-19*	19	37.82	37.61 - 38.44
34	LSZ	*qLSZ-19*	19	43.79	43.26 - 44.37
	GH	*qGH-19*	19	43.81	42.64 - 45.15
	LD	*qLD-19-1*	19	43.81	43.64 - 44.37
	LSH	*qLSH-19*	19	43.81	43.51 - 44.42
	SD	*qSD-19*	19	43.81	43.71 - 44.25
	SW	*qSW-19*	19	43.81	43.25 - 44.25
	R8	*qR8-19*	19	46.22	45.06 - 48.36
35	SW	*qSW-20*	20	32.45	28.10 - 33.47
	LSZ	*qLSZ-20*	20	34.50	34.11 - 34.90
36	R8	*qR8-20*	20	39.46	38.92 - 40.72

[†]R1DAP, first flowering (days after planting); R8, full maturity (days); SD, stem diameter (mm); MSL, main stem length (cm); SH, shattering (rating); GH, growth habit (rating); LD, lodging (rating); SW, 100-seed weight (g); LSZ, leaflet size (rating); LSH, leaflet shape (rating); PB, pubescence form (rating)

[‡]Chromosome

FIGURES

Figure 2.1. Density plots of phenotypic traits for both populations. Parental values indicated by vertical lines with PI 468916 in red, PI 479752 in blue, and Williams 82 in black.

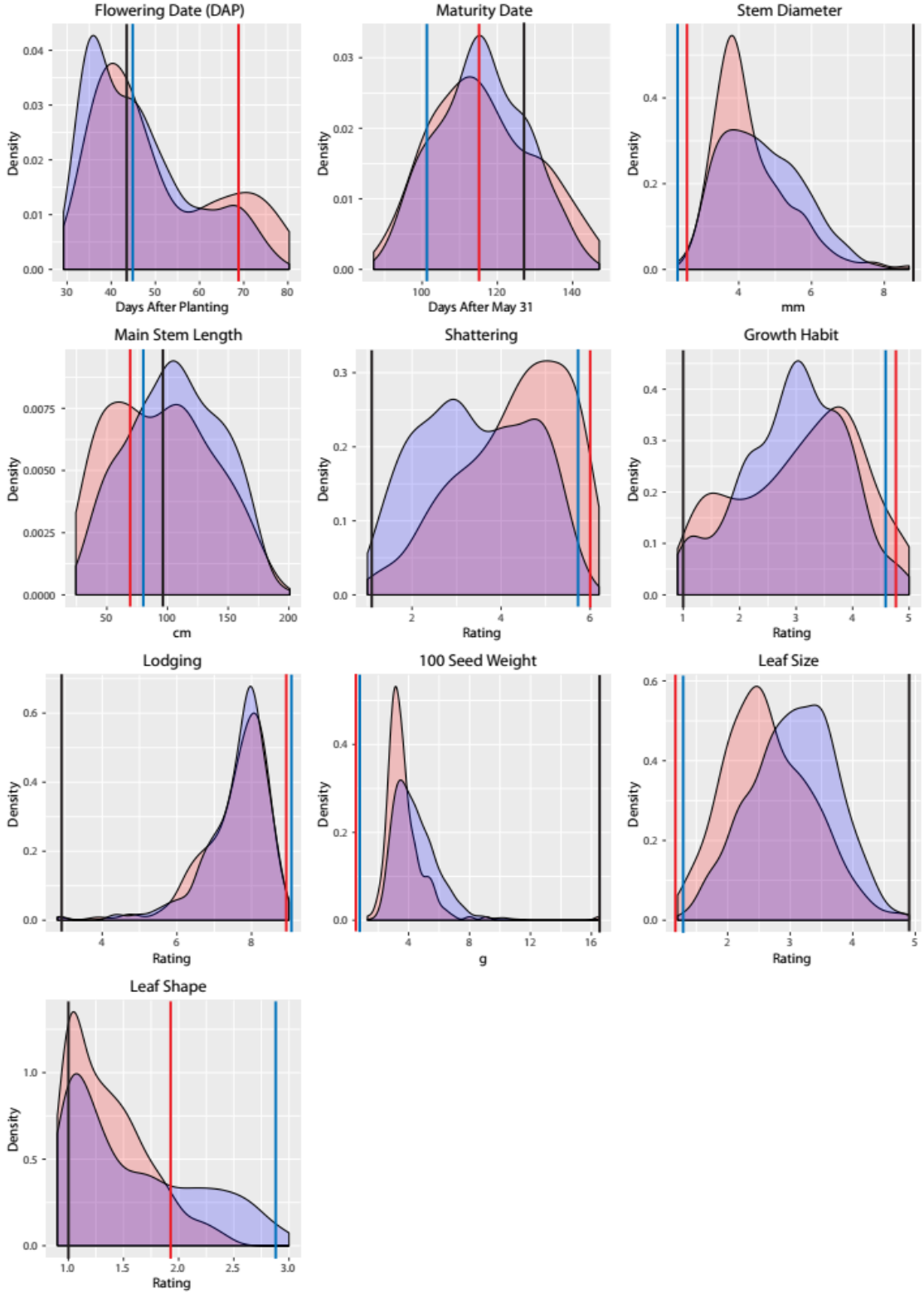
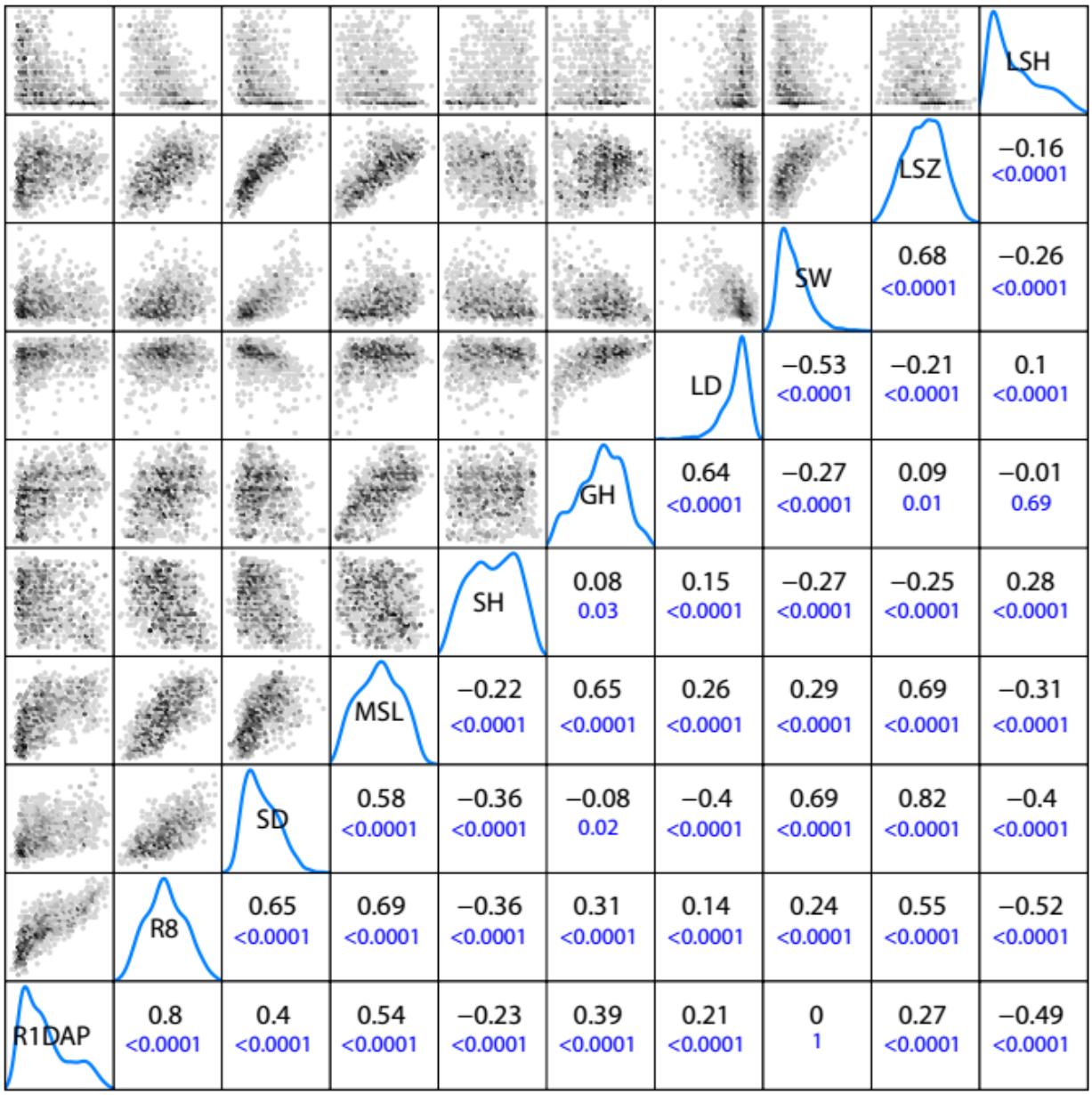


Figure 2.2. Among all 800 RILs, matrix of scatter plots (upper panel) and Pearson correlations (lower panel, with p-values in blue) of phenotypic traits among all RILs. Density distributions given on diagonal[†].



[†]R1DAP, first flowering (days after planting); R8, full maturity; SD, stem diameter; MSL, main stem length; SH, shattering; GH, growth habit; LD, lodging; SW, 100-seed weight; LSZ, leaflet size; LSH, leaflet shape

Figure 2.3. Genetic distance (cM) versus physical distance (Mb) of 7,673 SNPs polymorphic in both *G. max* x *G. soja* populations.

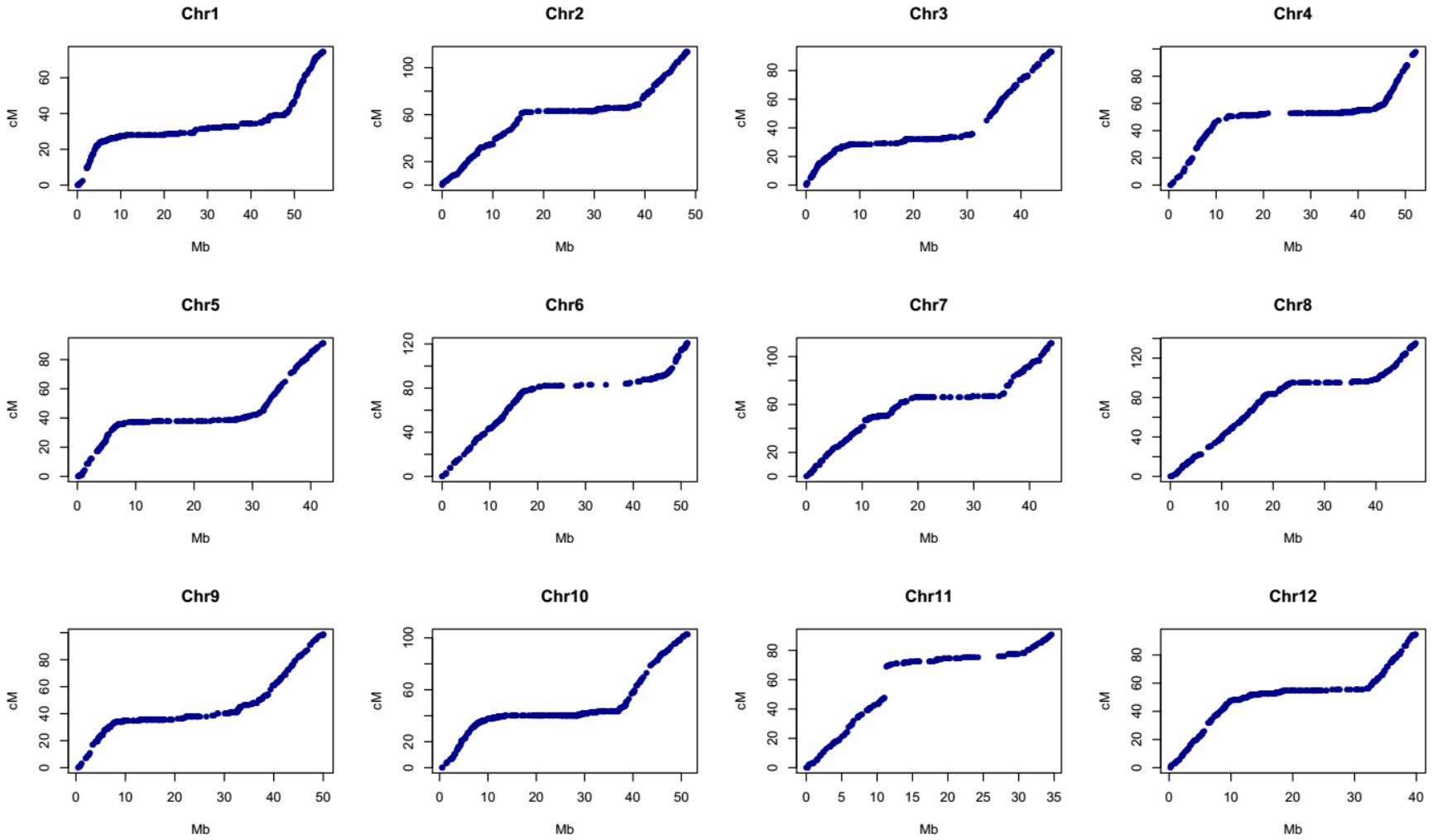


Figure 2.3 (cont.)

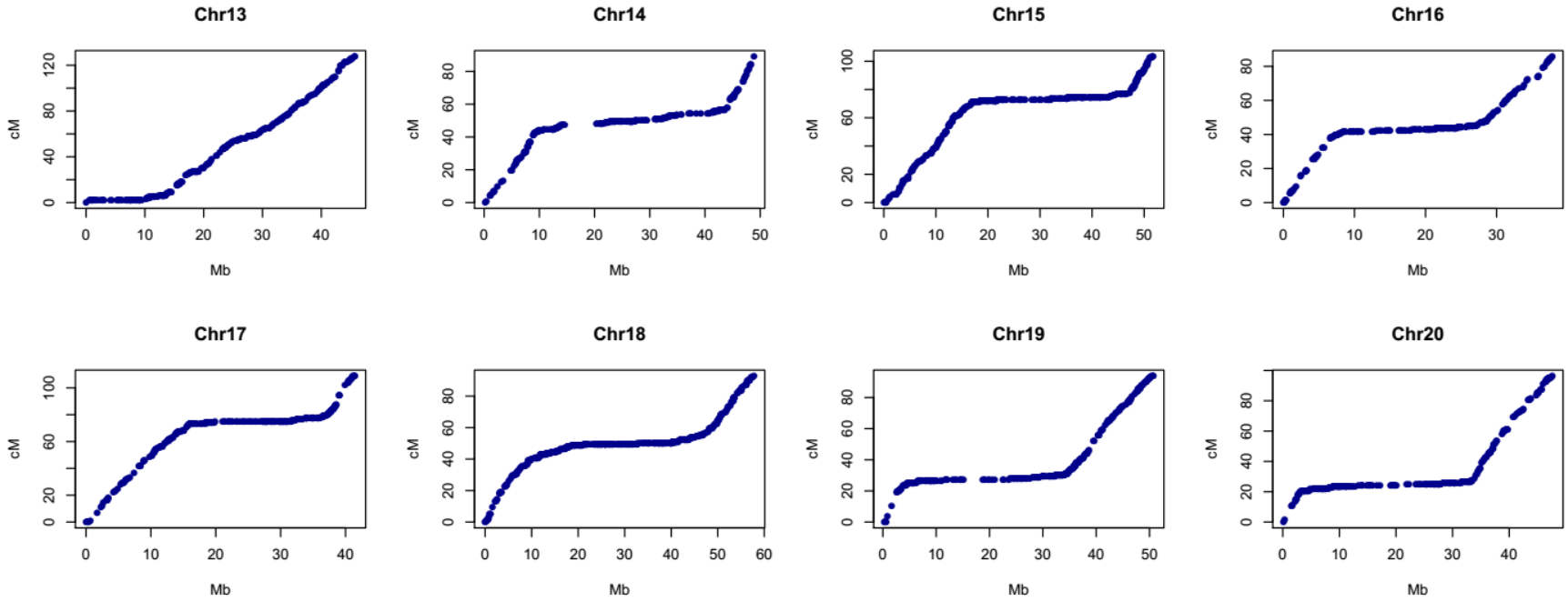


Figure 2.4. QTL on physical map as detected in Williams 82 x PI 468916. Horizontal lines indicate the location of the 16,893 SNP markers used for QTL mapping. Positions of the QTL are identified by colored circles. Chromosomal regions encompassed by the 1.5-LOD support intervals are highlighted in blue.

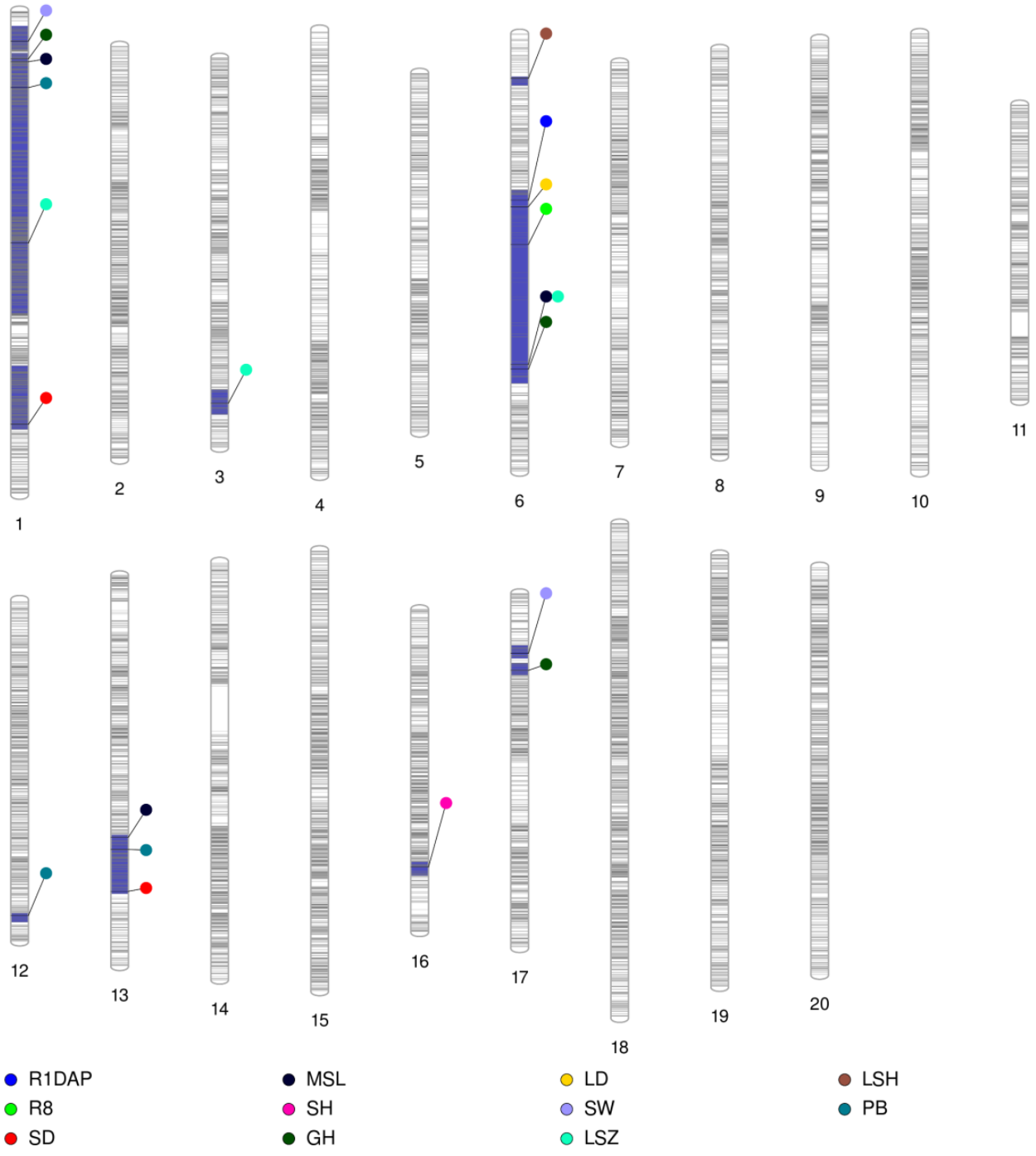


Figure 2.5. QTL on physical map as detected in Williams 82 x PI 479752. Horizontal lines indicate the location of the 17,052 SNP markers used for QTL mapping. Positions of the QTL are identified by colored circles. Chromosomal regions encompassed by the 1.5-LOD support intervals are highlighted in blue.

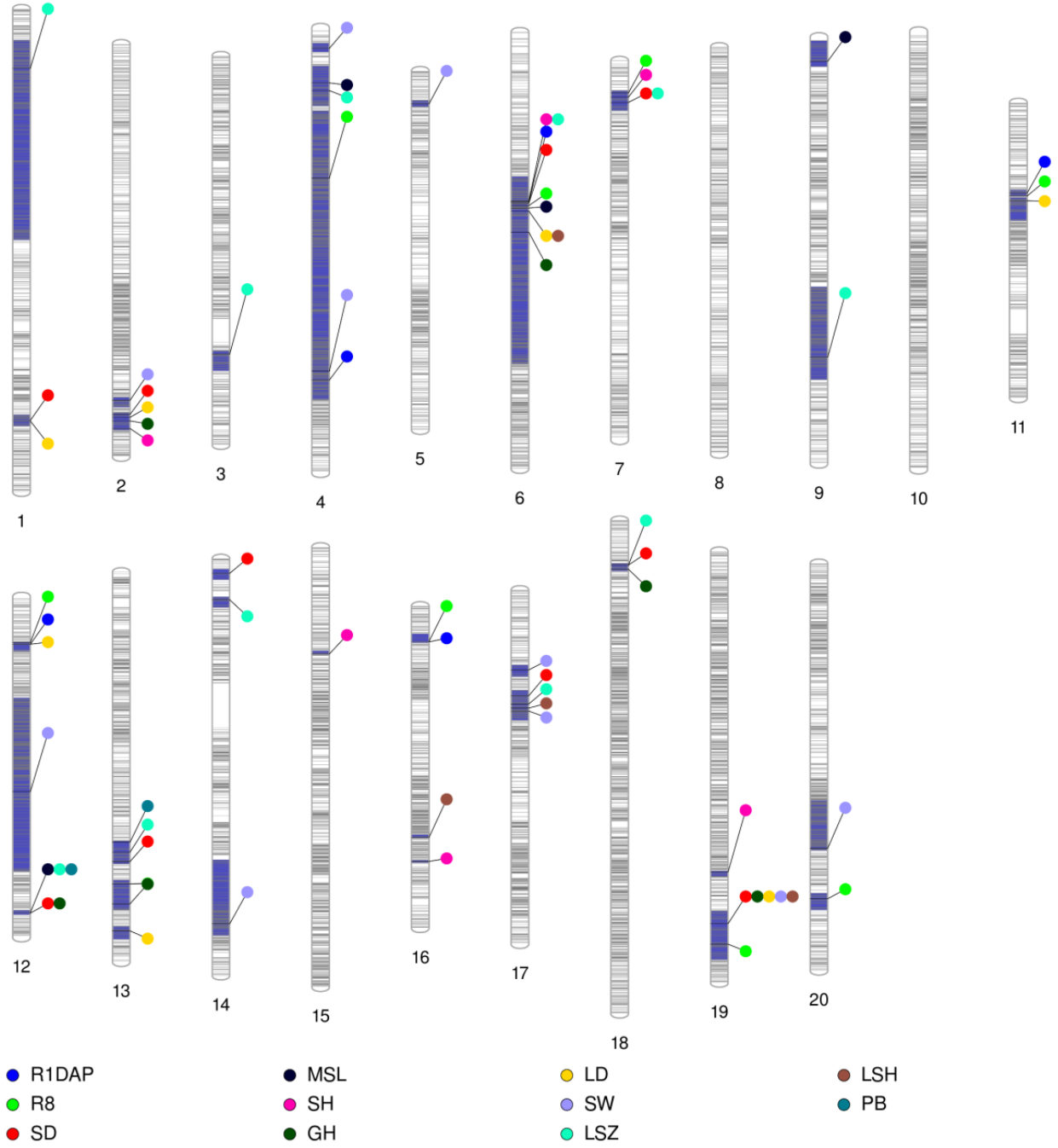
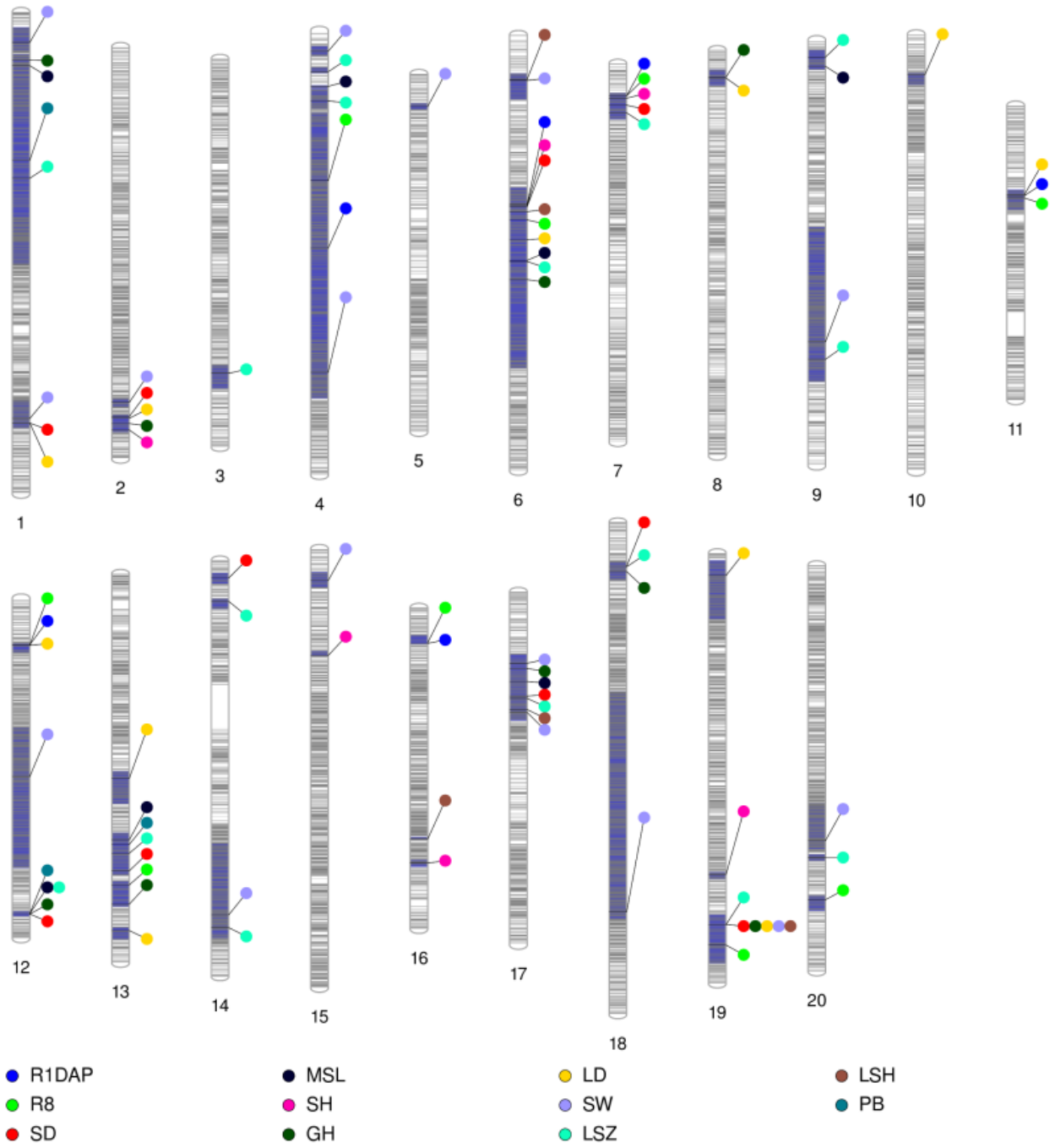


Figure 2.6. QTL on physical map as detected in Williams 82 x PI 468916, Williams 82 x PI 479752, and the combined populations. Horizontal lines indicate the location of the 26,272 SNP markers used for QTL mapping. Positions of the QTL are identified by colored circles, and chromosomal regions encompassed by the 1.5-LOD support intervals are highlighted in blue. The positions of QTL and support intervals were averaged from the three mapping results.



REFERENCES

Akpertey, A. 2015. Genetic introgression from *Glycine tomentella* to soybean to increase seed yield.

Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67(1): 1–48.

Beavis, W.D. 1998. QTL analyses: Power, precision, and accuracy. p. 145–162. In Paterson, A.H. (ed.), Molecular Dissection of Complex Traits. CRC Press, Boca Raton, FL.

Bernard, R.L. 1975. The inheritance of appressed pubescence. Soybean Genet. Newsl. 2: 34–36.

Broich, S.L., and R.G. Palmer. 1980. A cluster analysis of wild and domesticated soybean phenotypes. Euphytica 29(1): 23–32.

Broman, K.W., H. Wu, S. Sen, and G.A. Churchill. 2003. R/qtl: QTL mapping in experimental crosses. Bioinformatics 19(7): 889–890.

Carpenter, J.A., and W.R. Fehr. 1986. Genetic variability for desirable agronomic traits in populations containing *Glycine soja* germplasm. Crop Sci. 26(4): 681.

Carter, T.E., R.L. Nelson, C.H. Sneller, and Z. Cui. 2004. Genetic diversity in soybean. p. 303–416. In Boerma, H.R., Specht, J.E. (eds.), Soybeans: Improvement, production, and uses. American Society of Agronomy, Madison, WI.

Chen, H., Z. Shan, A. Sha, B. Wu, Z. Yang, S. Chen, R. Zhou, and X. Zhou. 2011. Quantitative trait loci analysis of stem strength and related traits in soybean. Euphytica 179(3): 485–497.

Chen, Q., Z. Zhang, C. Liu, D. Xin, H. Qiu, D. Shan, C. Shan, and G. Hu. 2007. QTL analysis of major agronomic traits in soybean. Agric. Sci. China 6(4): 399–405.

Cicek, M.S., P. Chen, M.A.S. Maroof, and G.R. Buss. 2006. Interrelationships among agronomic and seed quality traits in an interspecific soybean recombinant inbred population. Crop Sci. 46(3): 1253–1259.

Doebley, J.F., B.S. Gaut, and B.D. Smith. 2006. The molecular genetics of crop domestication. Cell 127(7): 1309–1321.

Doebley, J., and A. Stec. 1993. Inheritance of the morphological differences between maize and teosinte: Comparison of results for two F2 populations. Genetics 134(2): 559–570.

Doebley, J., A. Stec, and C. Gustus. 1995. *teosinte branched1* and the origin of maize: Evidence for epistasis and the evolution of dominance. Genetics 141(1): 333–346.

Dong, Y., X. Yang, J. Liu, B.H. Wang, B.L. Liu, and Y.Z. Wang. 2014. Pod shattering resistance associated with domestication is mediated by a NAC gene in soybean. Nat. Commun. 5: 3352.

Dorweiler, J., A. Stec, J. Kermicle, and J. Doebley. 1993. *Teosinte glume architecture 1*: A genetic locus controlling a key step in maize evolution. Science. 262(5131): 233–235.

Fehr, W.R., C.E. Caviness, D.T. Burmood, and J.S. Pennington. 1971. Stage of development descriptions for soybeans, *Glycine max* (L.) Merrill. Crop Sci. 11(6): 929–931.

Frary, A., T.C. Nesbitt, S. Grandillo, E. Knaap, B. Cong, J. Liu, J. Meller, R. Elber, K.B. Alpert, and S.D. Tanksley. 2000. *Fw2.2*: A quantitative trait locus key to the evolution of tomato fruit size. Science. 289(5476): 85–88.

Funatsuki, H., M. Hajika, T. Yamada, M. Suzuki, S. Hagihara, Y. Tanaka, S. Fujita, M. Ishimoto, and K. Fujino. 2012. Mapping and use of QTLs controlling pod dehiscence in soybean. Breed. Sci. 61(5): 554–558.

Funatsuki, H., M. Suzuki, A. Hirose, H. Inaba, T. Yamada, M. Hajika, K. Komatsu, T. Katayama, T. Sayama, M. Ishimoto, and K. Fujino. 2014. Molecular basis of a shattering resistance boosting global dissemination of soybean. Proc. Natl. Acad. Sci. 111(50): 17797–17802.

Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire, Q. Sun, and E.S. Buckler. 2014. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. PLoS One 9(2): e90346.

Grant, D., R.T. Nelson, S.B. Cannon, and R.C. Shoemaker. 2010. SoyBase, the USDA-ARS soybean genetics and genomics database.

Hammer, K. 1984. Das domestikationssyndrom. Die Kult. 32(3): 11–34.

Han, Y., X. Zhao, D. Liu, Y. Li, D.A. Lightfoot, Z. Yang, L. Zhao, G. Zhou, Z. Wang, L. Huang, Z. Zhang, L. Qiu, H. Zheng, and W. Li. 2016. Domestication footprints anchor genomic regions of agronomic importance in soybeans. New Phytol. 209(2): 871–884.

Higgins, R.H., C.S. Thurber, I. Assaranurak, and P.J. Brown. 2014. Multiparental mapping of plant height and flowering time QTL in partially isogenic sorghum families. G3 Genes|Genomes|Genetics 4(9): 1593–602.

Jeong, N., J.K. Moon, H.S. Kim, C.G. Kim, and S.C. Jeong. 2011. Fine genetic mapping of the genomic region controlling leaflet shape and number of seeds per pod in the soybean. Theor. Appl. Genet. 122(5): 865–874.

Jun, T.H., K. Freewalt, A.P. Michel, and R. Mian. 2014. Identification of novel QTL for leaf traits in soybean. Plant Breed. 133(1): 61–66.

Keim, P., B.W. Diers, T.C. Olson, and R.C. Shoemaker. 1990. RFLP mapping in soybean: Association between marker loci and variation in quantitative traits. Genetics 126(3): 735–742.

Kim, H.K., S.T. Kang, and D.Y. Suh. 2005. Analysis of quantitative trait loci associated with leaflet types in two recombinant inbred lines of soybean. Plant Breed. 124(6): 582–589.

Kong, F., B. Liu, Z. Xia, S. Sato, B.M. Kim, S. Watanabe, T. Yamada, S. Tabata, A. Kanazawa, K. Harada, and J. Abe. 2010. Two coordinately regulated homologs of FLOWERING LOCUS T are involved in the control of photoperiodic flowering in soybean. PLANT Physiol. 154(3): 1220–1231.

Kulkarni, K.P., M. Kim, J.G. Shannon, and J.D. Lee. 2016. Identification of quantitative trait loci controlling soybean seed weight in recombinant inbred lines derived from PI 483463 (*Glycine soja*) × "Hutcheson" (G. max). Plant Breed. 135(5): 614–620.

Kuroda, Y., A. Kaga, N. Tomooka, H. Yano, Y. Takada, S. Kato, and D. Vaughan. 2013. QTL affecting fitness of hybrids between wild and cultivated soybeans in experimental fields. Ecol. Evol. 3(7): 2150–2168.

Lam, H.M., X. Xu, X. Liu, W. Chen, G. Yang, F.L. Wong, M.W. Li, W. He, N. Qin, B. Wang, J. Li, M. Jian, J.J. Wang, G. Shao, J.J. Wang, S.S.M. Sun, and G. Zhang. 2010. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. Nat. Genet. 42(12): 1053–1059.

Langmead, B., C. Trapnell, M. Pop, and S.L. Salzberg. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10(3): R25.

Lee, J.M., A.L. Bush, J.E. Specht, and R.C. Shoemaker. 1999. Mapping of duplicate genes in soybean. Genome 42(5): 829–836.

Lenth, R. V. 2016. Least-squares means: The R package lsmeans. J. Stat. Softw. 69(1).

Li, Y., S. Zhao, J. Ma, D. Li, L. Yan, J. Li, X. Qi, X. Guo, L. Zhang, W. He, R. Chang, Q. Liang, Y. Guo, C. Ye, X. Wang, Y. Tao, R. Guan, J. Wang, Y. Liu, L. Jin, X. Zhang, Z. Liu, L. Zhang, J. Chen, K. Wang, R. Nielsen, R. Li, P. Chen, W. Li, J.C. Reif, M. Purugganan, J. Wang, M. Zhang, J. Wang, and L. Qiu. 2013. Molecular footprints of domestication and improvement in soybean revealed by whole genome re-sequencing. BMC Genomics 14(1): 579.

Li, C., A. Zhou, and T. Sang. 2006. Rice domestication by reducing shattering. Science. 311(5769): 1936–1939.

Liu, B., T. Fujita, Z.H. Yan, S. Sakamoto, D. Xu, and J. Abe. 2007. QTL mapping of domestication-related traits in soybean (*Glycine max*). Ann. Bot. 100(5): 1027–1038.

Mace, E.S., K.K. Buhariwalla, H.K. Buhariwalla, and J.H. Crouch. 2003. A high-throughput DNA extraction protocol for tropical molecular breeding programs. Plant Mol. Biol. Report. 21(4): 459–460.

Maughan, P.J., M.A.S. Maroof, and G.R. Buss. 1996. Molecular-marker analysis of seed-weight: genomic locations, gene action, and evidence for orthologous evolution among three legume species. TAG Theor. Appl. Genet. 93(4): 574–579.

Orf, J.H., K. Chase, T. Jarvik, L.M. Mansur, P.B. Cregan, F.R. Adler, and K.G. Lark. 1999. Genetics of soybean agronomic traits: Comparison of three related recombinant inbred populations. Crop Sci. 39(6): 1642.

Palmer, R.G., T.W. Pfeiffer, G.R. Buss, and T.C. Kilen. 2004. Qualitative genetics. p. 137–233. In Boerma, H.R., Specht, J.E. (eds.), Soybeans: Improvement, production, and uses. 3rd ed. American Society of Agronomy, Madison, WI.

Peng, J., Y. Ronin, T. Fahima, M.S. Roder, Y. Li, E. Nevo, and A. Korol. 2003. Domestication quantitative trait loci in *Triticum dicoccoides*, the progenitor of wheat. Proc. Natl. Acad. Sci. 100(5): 2489–2494.

R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Simons, K.J. 2005. Molecular characterization of the major wheat domestication gene *Q*. Genetics 172(1): 547–555.

Singh, R.J., and T. Hymowitz. 1988. The genomic relationship between *Glycine max* (L.) Merr. and *G. soja* Sieb. and Zucc. as revealed by pachytene chromosome analysis. Theor. Appl. Genet. 76(5): 705–711.

Song, Q., D.L. Hyten, G. Jia, C. V. Quigley, E.W. Fickus, R.L. Nelson, and P.B. Cregan. 2013. Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. PLoS One 8(1): 1–12.

Song, Q., J. Jenkins, G. Jia, D.L. Hyten, V. Pantalone, S.A. Jackson, J. Schmutz, and P.B. Cregan. 2016. Construction of high resolution genetic linkage maps to improve the soybean genome sequence assembly Glyma1.01. BMC Genomics 17(1): 33.

Sugimoto, K., Y. Takeuchi, K. Ebana, A. Miyao, H. Hirochika, N. Hara, K. Ishiyama, M. Kobayashi, Y. Ban, T. Hattori, and M. Yano. 2010. Molecular cloning of *Sdr4*, a regulator involved in seed dormancy and domestication of rice. Proc. Natl. Acad. Sci. 107(13): 5792–5797.

Sun, L., Z. Miao, C. Cai, D. Zhang, M. Zhao, Y. Wu, X. Zhang, S.A. Swarm, L. Zhou, Z.J. Zhang, R.L. Nelson, and J. Ma. 2015. *GmHs1-1*, encoding a calcineurin-like protein, controls hard-seededness in soybean. Nat. Genet. 47(8).

Swarts, K., H. Li, J.A. Romero Navarro, D. An, M.C. Romay, S. Hearne, C. Acharya, J.C. Glaubitz, S. Mitchell, R.J. Elshire, E.S. Buckler, and P.J. Bradbury. 2014. Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants. Plant Genome 7(3): 0.

Thurber, C.S., J.M. Ma, R.H. Higgins, and P.J. Brown. 2013. Retrospective genomic analysis of sorghum adaptation to temperate-zone grain production. Genome Biol. 14(6): R68.

Tian, Z.X., X.B. Wang, R. Lee, Y.H. Li, J.E. Specht, R.L. Nelson, P.E. McClean, L.J. Qiu, and J.X. Ma. 2010. Artificial selection for determinate growth habit in soybean. Proc. Natl. Acad. Sci. 107(19): 8563–8568.

Wang, W., Q. He, H. Yang, S. Xiang, T. Zhao, and J. Gai. 2013. Development of a chromosome segment substitution line population with wild soybean (*Glycine soja* Sieb. et Zucc.) as donor parent. Euphytica 189(2): 293–307.

Wolfe, D., S. Dudek, M.D. Ritchie, and S.A. Pendergrass. 2013. Visualizing genomic information across chromosomes with PhenoGram. BioData Min. 6(1): 18.

Xia, Z., S. Watanabe, T. Yamada, Y. Tsubokura, H. Nakashima, H. Zhai, T. Anai, S. Sato, T. Yamazaki, S. Lu, H. Wu, S. Tabata, and K. Harada. 2012. Positional cloning and characterization reveal the molecular basis for soybean maturity locus E1 that regulates photoperiodic flowering. Proc. Natl. Acad. Sci. 109(32): E2155–E2164.

Xin, D., Z. Qi, H. Jiang, Z. Hu, R. Zhu, J. Hu, H. Han, G. Hu, C. Liu, and Q. Chen. 2016. QTL location and epistatic effect analysis of 100-seed weight using wild soybean (*Glycine soja* Sieb. & Zucc.) chromosome segment substitution lines. PLoS One 11(3): e0149380.

Yan, L., Y.H. Li, C.Y. Yang, S.X. Ren, R.Z. Chang, M.C. Zhang, and L.J. Qiu. 2014. Identification and validation of an over-dominant QTL controlling soybean seed weight using populations derived from *Glycine max* × *Glycine soja*. Plant Breed. 133(5): 632–637.

Zhou, Z., Y. Jiang, Z. Wang, Z. Gou, J. Lyu, W. Li, Y. Yu, L. Shu, Y. Zhao, Y. Ma, C. Fang, Y. Shen, T. Liu, C. Li, Q. Li, M. Wu, M. Wang, Y. Wu, Y. Dong, W. Wan, X. Wang, Z. Ding, Y. Gao, H. Xiang, B. Zhu, S.H. Lee, W. Wang, and Z. Tian. 2015. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. Nat. Biotechnol. 33(4): 408–14.

CHAPTER THREE

Evaluation of *G. soja*-Derived Lines

INTRODUCTION

The genetic base of the current North American soybean genepool is limited in scope. Gizlice et al. (1994) surveyed the pedigrees of public varieties released in North America between 1947 and 1988. They found that just 80 ancestral lines could account for the entire genetic base of these varieties. In addition, only six of these ancestors constitute over half of the genetic base. This underlines the need to broaden the germplasm pool of soybean breeding.

A few different germplasm sources are available to soybean breeders to expand the genetic base, namely exotic *G. max* accessions, *G. soja*, and members of the subgenus *Glycine*. Landraces of soybean would generally require the least breeding effort to adjust the adaptation to North American environments. However, when comparing the genetic diversity of landrace populations with genotypes from modern breeding programs, landraces typically only contain a slightly higher level of diversity (Hyten et al., 2006; Zhou et al., 2015). The perennial relatives of soybean in the subgenus *Glycine* are quite diverged from *G. max* and would likely represent a significant source of novel genes and alleles. The genetic differences are large enough that hybridizations require embryo rescue and are usually unsuccessful, with only a few exceptions (Newell et al., 1987; Singh and Nelson, 2015). On the other hand, annual wild soybean, *Glycine soja*, easily hybridizes with *G. max* to produce fertile progeny. *G. soja* also contains twice the genetic diversity found in soybean cultivars (Hyten et al., 2006; Zhou et al., 2015).

The diversity present in wild soybean and the ease of crossing has led some breeders to utilize *G. soja* for improving a variety of traits. Crop wild relatives have most often been used to introgress genes for disease and pest resistance or tolerance to abiotic stresses (Hodgkin and Hajjar, 2008). *G. soja* has been shown to contain genes that can improve cold tolerance (Yang et al., 2014), drought tolerance (Luo et al., 2013; Tang et al., 2013), and SCN resistance (Kim and Diers, 2013). Several studies have also researched wild soybean's utility to improve yield. Concibido et al. (2003) reported a QTL from PI 407305 that led to a 9.4% yield increase in the mapping population. When introduced into additional genetic backgrounds, the QTL was associated with a significant yield effect in two out of six populations. The lack of a significant

effect in two thirds of the genetic backgrounds may be attributed to either the low number of environments, epistatic interactions with other loci, or the presence of the same allele in those *G. max* backgrounds. Li et al. (2008) also identified a QTL for which the *G. soja* allele from PI 245331 contributed a 6.3% increase in yield over the *G. max* allele. While not verified in other *G. max* backgrounds, the QTL was validated using a separate population. These QTL mapping studies have demonstrated the presence of positive alleles in wild soybean and the potential for improving soybean.

While *G. soja* is known to contain beneficial alleles and genes, applied uses of the germplasm in breeding programs is difficult. Detrimental traits such as lodging, shattering, and small seeds persist in the populations and typically require multiple backcrosses to *G. max* to detect agronomically acceptable progeny (Carpenter and Fehr, 1986). Examples of its limited use include a small-seeded specialty variety (Cui et al., 2004), a germplasm line with SCN resistance (Diers et al., 2005), and two germplasm lines with 50% pedigree from wild soybean (Abdel-Haleem et al., 2015). A previous study by Akpertey et al. (2014) evaluated 93 lines derived from backcrosses between five *G. soja* accessions and four soybean recurrent parent backgrounds (either Williams 82 or Williams 82 and another cultivar). While they were unable to identify lines that were significantly higher yielding than the best soybean parent, they did find lines that were equivalent in yield at α=0.05. They also found that the proportion of alleles derived from *G. soja* did not deviate from the expected ratios in the absence of selection.

The objectives of this current study were to identify high-yielding *G. soja*-derived lines and to characterize patterns of introgression from the wild accessions, especially at QTL for domestication-related traits (DRT) and regions of selective sweeps in *G. max*.

MATERIALS AND METHODS

Plant Materials

This study utilized 416 existing *G. soja*-derived lines from the USDA soybean breeding program at Urbana, IL. Therefore, the evaluated lines were derived from a diverse set of pedigrees. A total of 26 different crosses were represented (Table 3.1), involving 14 different *G. soja* accessions (Table 3.2) and 11 *G. max* parents. As shown in Table 3.1, the pedigree contribution for each *G. soja* parent varied from 6.25% to 50%. While a backcross is defined as

multiple crosses to the same genotype (recurrent parent), a liberal definition is applied here as multiple crosses to *G. max* regardless of genotype. All lines were heavily selected for agronomic traits during the inbreeding process to the F_5 generation or later. Three lines in this evaluation originated from the study by Akpertey et al. (2014), LG08-4152, LG07-4231, and LG08-3465. Other lines were related to the lines evaluated by Akpertey et al. (2014) study—03JR309156, 06NB204846, IA2052, and IA3023 were crossed to lines that originated from the same *G. max* x *G. soja* populations as the previous study.

Experimental Design

In 2015, 416 lines were divided into three tests based on maturity with 101 lines in MS II, 230 in MS III, and 90 in MS IV. All three tests were grown at locations in Urbana, IL and Ivesdale, IL. Each test was randomized as an $\alpha(0,1,2)$ design with two replications with 19 to 36 entries per block. Due to deficiencies in seed, 172 lines were grown in one replication at two locations. All available soybean parents were randomized as entries with two replications, and a check cultivar augmented each block. Plots were four rows wide with 0.76 m spacing and 3.1 m in length. Seeds were sown at 30 seeds m^{-1} on May 7 at Urbana and May 6 at Ivesdale. Data were collected on maturity date at R8 growth stage (95% of pods reached final color), height (measured as the distance from the soil surface to the top of the main stem), lodging (rated to nearest 0.5 score on scale from 1 to 5, with 1 completely erect and 5 prostrate), seed yield measured at 13% moisture, and 100-seed weight. At the Urbana location, shattering (percentage of pods dehisced two weeks after maturity, estimated to nearest 5%) and growth habit (visual rating to nearest 0.5 score on a 1 to 5 scale, with 1 representing typical indeterminate soybean and 5 representing vining growth similar to *G. soja*) were also measured. These data were used to select 300 lines for further testing by removing heavily shattered lines and poorly represented pedigrees (Tale 3.1).

In 2016, the 300 lines were again assigned to three tests based on their maturity. Each test was grown at three locations in Illinois, and one location in Nebraska and Missouri. MS II was planted at Urbana, IL (May 23), Ivesdale, IL (May 19), Pontiac, IL (May 6), Columbia, MO (May 23), and Mead, NE (June 3). MS III was planted at Urbana, IL (May 23), Ivesdale, IL (May 19), Arthur, IL (May 18), Columbia, MO (May 23), and Mead, NE (June 2). MS IV was planted at Urbana, IL (May 23), Ivesdale, IL (May 19), Neoga, IL (May 25), Columbia, MO (May 23), and Clay Center, NE (May 20). An $\alpha(0,1,2)$ design with two replications for all lines

was used, with blocks containing 22 or 26 entries. Maturity date (R8), height, lodging, and seed yield were collected at all locations. At Urbana and Ivesdale, flowering date (50% of plants reached R1 with at least one flower), shattering, and growth habit were also measured.

DNA Isolation and Genotyping-by-Sequencing

For each line, tissue was collected from 10 plants grown in a greenhouse. Newly-expanded leaf tissue was sampled and bulked for DNA extraction using a modified CTAB protocol (Mace et al., 2003). The GBS library was prepared following a modification of the two-enzyme protocol described in Thurber et al. (2013). Approximately 250 ng of DNA was digested with *HindIII* and *MseI*, which was then ligated with one of 768 unique barcoded adapters with an *HindIII* overhang. The library of 466 samples was sequenced at the W. M. Keck Center at the University of Illinois using an Illumina HiSeq4000 with 100-bp single-end reads, resulting in an average of over 500 thousand reads per sample.

Processing of the Illumina reads was performed in TASSEL 5.0 using the GBSv2 pipeline (Glaubitz et al., 2014). Default settings were used for most parameters and the maximum tag length was set to 80 bp. Sequenced tags were aligned to the Williams 82 reference genome (Wm82.a2.v1 downloaded from Phytozome) using Bowtie 2 with the "very-sensitive" option (Langmead et al., 2009). Indels and SNPs with inbreeding coefficients (F) less than 0.7 or minor allele frequencies less than 0.01 were removed. Before filtering, over 99 thousand SNPs were called, and after removing markers with greater than 50% missing data and minor allele frequency less than 1%, 57,222 SNPs remained. Beagle was used to impute missing data with a window size of 500 SNPs and overlap of 100 SNPs (Browning and Browning, 2016). Parental alleles were called for crosses in which all parents were genotyped. Only markers homozygous and polymorphic in the parents were used. Alleles matching the *G. soja* parent were coded as '2', alleles matching the *G. max* parent(s) were coded as '0', and heterozygous loci were '1'.

Data Analysis

Phenotypic traits were analyzed in R using the 'lme4' package for mixed models (Bates et al., 2015). Genotypes were treated as fixed effects, while environment (year by location), replication nested within environment, block nested within replication, and genotype by environment interaction were treated as random effects. Least-squares means of the genotypes were calculated using the 'lsmeans' package in R (Lenth, 2016). Mean separation of genotype

means was performed with Fisher's protected LSD using the residual variance calculated by lme4.

RESULTS

Agronomic Performance

Tables 3.3, 3.4, and 3.5 present the means of agronomic traits for each entry across all environments for tests MS II, MS III, and MS IV, respectively. The three tests were established to roughly correlate to maturity groups II, III, and IV. In each test, the *G. soja*-derived entries were generally slightly later in maturity than the corresponding yield check, (average +4, +4, and +3 days relative to the check in MS II, MS III, and MS IV respectively). Significant differences among the genotypes were detected for all eight traits. Environmental effects and genotype by environment interactions were also significant ($\alpha = 0.05$) for the majority of traits. No environmental differences were detected for growth habit in MS II, shattering in MS II and MS IV, or 100-seed weight in MS III, while all traits contained significant genotype by environment interactions.

The *G. soja*-derived lines displayed a range for several agronomic related traits, including plant height, lodging, growth habit, shattering, and 100-seed weight. Plant height ranged from 58 cm (LG14-3458 in MS II) to 150 cm (LG14-3000 in MS IV). This distribution far exceeded the range in plant heights measured among soybean parents and checks which fell within 99 cm to 128 cm. While some lines displayed lodging scores similar to the soybean checks (e.g. LG14-2458 at 1.5), the large majority of *G. soja*-derived lines were more lodged than the cultivars. A total of 310 lines had lodging scores greater than LG11-6210 (*G. max* check replicated across tests in 2016). Growth habit also varied among lines with scores ranging between 1.0 and 2.5. The majority of lines displayed growth habits comparable to typical soybean cultivars with 314 lines averaging growth habit scores of 1.5 or less, and only 27 lines displayed slight vining tendencies with growth habit 2.0 or over. *G. soja*-derived lines also were predominantly shattering resistant with 337 averaging 0% shattering across environments. Thirteen lines displayed degrees of shattering susceptibility with 15 lines that averaged over 25% shattered pods and two lines (LG14-3169 and LG14-3835) that averaged over 40%. The material in this study also displayed a considerable range in 100-seed weights, with the smallest seeds (LG14-

2756) weighing only 5.1 g and the largest (LG14-7933) weighing 20.3 g. A total of 66 lines had significantly larger 100-seed weights compared to the corresponding yield check for the test, with 64 lines derived from pedigrees containing 03JR309156 or 06NB204846 as parents.

For each test, the highest yielding line was derived from crosses with the Syngenta soybean cultivar 03JR309156 (LG14-6052 with 4922 kg/ha in MS II, LG14-6099 with 5295 kg/ha in MS III, and LG14-7948 with 4765 kg/ha in MS IV). Lines derived from 06NB204846 (another Syngenta cultivar) were also among the highest in yields across pedigrees. When comparing the *G. soja*-derived lines to the yield checks for each test (LD02-4485 in MS II, LD08-8622 in MS III, and LD06-7620 in MS IV), only two lines in MS II (LG14-6052 and LG14-5100) were significantly higher in yield ($\alpha = 0.05$) and both were derived from one of the two Syngenta parents. A portion of this yield difference could potentially be explained by a difference in reproductive period, as LG14-6052 and LG14-5100 were approximately a week later in maturity relative to LD02-4485. The cross of LN97-15076 x LG01-7770 was also generally high-yielding, with the fourth highest average yield among crosses (Table 3.6). Also shown in Table 3.6, crosses with 50% contribution from *G. soja* yielded less than crosses with backcrosses to *G. max*. With the exception of Macon x PI 483461 (represented by only 2 lines), all pedigrees without backcrosses averaged under 3500 kg/ha. The lowest yielding cross was IA3023 x PI 522183A which averaged only 2446 kg/ha. Most crosses displayed considerable ranges for yield (Table 3.6). Backcross pedigrees tended to display smaller ranges in yield (under 1100 kg/ha), while the straight-cross pedigrees generally displayed ranges over 1200 kg/ha, especially Williams 82 x PI 479752 which encompassed a 3558 kg/ha range.

Due to a lack of available seed source, only a portion of the *G. max* genotypes involved in the pedigrees were able to be included as entries in the tests, namely LN97-15076, Williams 82, SS98-3403, LG04-6000, and IA3023. All lines derived from IA3023, LG04-6000, and SS98-3403 were significantly lower yielding than their *G. max* parent. In the LN97-15076 x LG01-7770 pedigree, two lines from MS IV (LG14-3897 and LG14-3899) were numerically higher in yield than LN97-15076, but were not significantly greater at $\alpha = 0.05$ (Table 3.7). As a maturity group IV cultivar, LN97-15076 was only included as an entry to the MS IV test. However, five lines from that pedigree were also grown in MS III. Only one of these lines, LG14-3902, had an average yield numerically lower than LN97-15076, but LG14-3894 yielded 387 kg/ha (109%) more than LN97-15076. Regrettably, since LN97-15076 was not grown in the MS III test,

comparisons to the soybean parent across tests should cautiously be made given the presence of environment and genotype by environment effects. Nearly a third of the 416 lines tested in this study were derived from crosses with Williams 82, with 124 lines originating from Williams 82 x PI 479752. Williams 82 was included as an entry in MS III, which makes comparisons within that test the most accurate. Lines from MS III that were statistically equivalent or greater in yield to Williams 82 are presented in Table 3.8. In the MS III test, all three Williams 82 (2) x PI 479767 progeny were statistically equivalent in yield to the soybean parent at $\alpha = 0.05$. In the Williams 82 x 479752 pedigree, 16 out of 53 lines were statistically equivalent to Williams 82, and two lines (LG14-3220 and LG12-9591) were significantly higher in yield at 109% of Williams 82 and while also equivalent in maturity.

G. soja Contributions and SNP Distribution

The genetic contribution from *G. soja* was calculated for lines for which all parents were genotyped. Table 3.9 presents results from the total contribution from *G. soja* in these crosses. As expected, lines that were backcrossed to *G. max* (BC_1 and BC_2) contained fewer alleles derived from *G. soja* than lines from non-backcrosses. For all crosses, the average percentage of *G. soja* alleles was lower than what would be expected in a population without selection. The differences were significant for all crosses except in the two BC_2 pedigrees (LN97-15076 x LG01-7770 and SS98-3403 (3) x PI 483461). LG04-6000 x PI 549048 had the highest overall *G. soja* allele frequency at 26.9% averaged across 43 lines. (Table 3.9). The line with the largest contribution from *G. soja* (41.7%) originated Williams 82 x PI 479752. This pedigree also represented the largest range, with lines containing as little as 8.8% of alleles from *G. soja*.

Due to the large number of lines in Williams 82 x PI 479752, these lines were used to calculate the *G. soja* allele frequency distribution across each chromosome (Fig. 3.1). Allele frequencies were calculated using a sliding window of 20 SNPs to smooth the distribution. As reflected by the average total *G. soja* allele frequency described in Table 3.9 (25.8%), the majority of chromosome regions were below the expected frequency of 50%. Some regions contained *G. soja* alleles at frequencies over 50% (e.g. sections of chr 3, 4, and 18), while other regions were nearly fixed for the *G. max* allele (e.g. regions of chr 2, 13, and 17). Only 10% of the SNP windows had greater than 50% *G. soja* allele frequency.

DISCUSSION AND CONCLUSION

Appearance of *G. soja*-Derived Lines

As described by Carpenter and Fehr (1986), progeny in the early generations from *G. max* x *G. soja* populations lack desirable agronomic traits, especially lodging resistance and absence of vining. The recommendation from their study is to use multiple backcrosses to *G. max* when selection is not employed. The lines evaluated in this study were subjected to stringent selection for agronomic appearance traits, including lodging, growth habit, and shattering. By imposing this selection pressure during the inbreeding process, the lines from this study largely recovered the appearance of the soybean parents. While Carpenter and Fehr (1986) were only able to recover the agronomic score of the soybean recurrent parent with two backcrosses or more, lines with 50% *G. soja* by pedigree were identified that were equivalent to the soybean parent for agronomic traits in this study. This demonstrates the ability to retain acceptable agronomic appearance with limited or no backcrossing if sufficient selection is imposed on large populations. Growth habit and shattering were efficiently improved through selection, as illustrated by the small number of lines with undesirable levels for the traits (i.e. 27 lines with growth habit ratings 2 or greater, and 15 lines with shattering significantly greater than 0%). Undesirable levels of lodging and seed weight were more frequently observed, where 74 lines with lodging scores 4 or above and 206 lines with 100-seed weight below 10 g.

Time of Flowering

Some pedigrees displayed considerable ranges in flowering date. Lines from IA3023 x PI 522183A, Williams 82 x PI 479752, and LG04-6000 x PI 549048 ranged over 30 days between the earliest and latest flowering. Although flowering date and maturity date exhibited a positive correlation, there was still substantial variation in their relationship. From the cross of IA3023 x PI468397A, LG14-2607 and LG14-2573 only matured one day apart, but LG14-2607 flowered ten days earlier than LG14-2573 (Table 3.4). A more extreme example can be seen within LG04-6000 x PI 549048. LG14-3557 matured on the same date as LG14-3578, yet flowered 20 days earlier (Table 3.4). Within Williams 82 x PI 479752, LG14-3378 matured two days later than LG14-3323 but initiated flowering 25 days sooner (Table 3.5).

Flowering was also not always correlated with yield, as some late flowering lines yielded the same as or higher than those that flowered early. LG14-3439 flowered 10 days later than

LG14-3261 (while maturing on the same day), yet yielded 847 kg/ha more (Table 3.4). In the Williams 82 x PI 479752 cross, LG14-3176 and LG14-3287 both matured 125 days after May 31 (Table 3.5), but LG14-3176 flowered 11 days after LG14-3287 yet yielded 838 kg/ha more. Such comparisons indicate that a longer period between flowering and maturity does not necessarily translate to gains in yield. The time between R5 and R8 is the more critical period for determining seed yield (Nelson, 1986) so without R5 dates we don't know the actual length of the seed filling period. Additional research on these lines could determine if those lines with late flowering and high yield have higher seed dry weight accumulation rates and if that trait could be combined with a longer seed filling period to achieve even higher yields.

Yield of *G. soja*-Derived Lines Compared to Soybean Parents

The utility of *G. soja* for soybean yield improvement can best be judged by comparing the performance of the progeny to the soybean parent. Lines derived from crosses with 03JR309156 and 06NB204846 were among the highest yielding in this study and some were equivalent to or exceeded the yield of soybean checks (Tables 3.3, 3.4, and 3.6). Regrettably, 03JR309156 and 06NB204846 were not available to be used in this study and therefore the performance of these top-yielding lines relative to their *G. max* parent cannot be made. It is therefore undetermined whether the higher yields of these lines are due to QTL from the *G. soja* accessions or simply due to higher yielding soybean parents used in the crosses.

Such comparisons are instead limited to progeny from crosses with five different *G. max* parents—IA3023, LG04-6000, SS98-3403, LN97-15076, and Williams 82. Among the crosses involving LN97-15076 and Williams 82, 24 lines were identified that were equivalent in yield to the soybean parent. LN97-15076 was the highest yielding soybean parent in the cross of LN97-15076 x LG01-7770 and had the largest contribution by pedigree (Table 3.1). As shown in Table 3.5, two lines derived from this cross, LG14-3897 and LG14-3899, were tested in the same environments as LN97-15076 and yielded 4539 kg/ha and 4533 kg/ha respectively (104% of the yield of LN97-15076). These yields are above the soybean parent, but the differences are not significant at α=0.05. These two lines were similar to LN97-15076 for the other traits as well, although a few differences were observed. LG14-3897 was 5 days later in maturity, and both were lower in 100-seed weight (-2.4 g and -1.3 g).

The large majority of lines that recovered the yield of the soybean parent were derived from Williams 82. In MS III, three pedigrees contained Williams 82 as the sole *G. max* parent—

Williams 82 x LG01-7909, Williams 82 (2) x PI 479767, and Williams 82 x PI 479752. *G. soja* contributed 12.5% by pedigree to LG07-4231 (Williams 82 x LG01-7909). At $\alpha = 0.05$, this line was not significantly different from the yield of Williams 82, but it did yield 189 kg/ha more while maturing one day earlier. Three other backcross lines (from Williams 82 (2) x PI 479767, 25% *G. soja* by pedigree) were also equivalent in yield to Williams 82, while maturing three to five days earlier. Surprisingly, out of 53 Williams 82 x PI 479752 lines (50% *G. soja* by pedigree), 16 were equivalent in yield to Williams 82 and 2 were significantly greater in yield at $\alpha = 0.05$ (Table 3.8). LG14-3220 and LG12-9591 yielded 343 kg/ha and 338 kg/ha more than Williams 82, respectively. Both lines were similar to Williams 82 for other agronomic traits. LG14-3220 displayed a slight increase in lodging (3.5 compared to 2.9), but was equivalent to Williams 82 in flowering date, maturity, height, and shattering. LG12-9591 was shorter than Williams 82, which likely contributed to its improved lodging score (2.5 compared to 2.9). Shattering (0%) and growth habit (1.0) were also equivalent to Williams 82. However, both LG14-3220 and LG12-9591 had significantly lower 100-seed weight (14.3 and 12.5 g, respectively) than Williams 82 (16.1 g), which indicates the increase in yield was due to an increase in the number of seeds, as opposed to an increase in seed size.

No lines from IA3023, LG04-6000, or SS98-3403 recovered or exceeded the yield of the soybean parent. The majority of lines derived from IA3023 and LG04-6000 contained 50% *G. soja* by pedigree. With such high contribution from the wild parent, it is not surprising that we were unable to regain the yields of the soybean parent (although this was not the case in the Williams 82 x PI 479752 pedigree). Recently two germplasm releases were made by the University of Georgia Agricultural Experiment Stations, both of which have 50% *G. soja* contributions by pedigree (Abdel-Haleem et al., 2015). While these germplasm lines were well above the average yields of *G. max* x *G. soja* populations, they were still only 88% of the yield of the soybean parent. While these lines would be useful germplasm for breeders to bridge the two species, this illustrates the difficulty in improving upon the yield of the soybean parent when *G. soja* contributes a large proportion of the pedigree. We were unable to regain the yield of SS98-3403 even with two backcrosses. SS98-3403 and five lines from the SS98-3403 (3) x PI 483461 pedigree were grown as entries in MS III in 2016. All of these lines were significantly lower yielding than the *G. max* parent with yields of 79% to 91% of SS98-3403 and while also maturing later (Table 3.4). All five lines had 100-seed weights approximately 7 g lower than

SS98-3403, which could contribute to the lower yields. This demonstrates that in *G. max* x *G. soja* crosses, recovery or improvement of the soybean parent yield is not guaranteed in spite of multiple backcrosses.

The inability to recover the yield of IA3023, LG04-6000, or SS98-3403 underscores the significance of developing high-yielding lines derived from Williams 82 and LN97-15076. Recovering the yield of the soybean parent using *G. max* x *G. soja* crosses is notable, and the occurrence of two lines that are significantly higher yielding is a considerable achievement. Interestingly, the success of *G. max* x *G. soja* crosses appeared to be inversely related to the yield of the soybean parent. Williams 82 and LN97-15076 were among the lowest yielding parents used in this study at 3954 kg/ha and 4380 kg/ha respectively (Tables 3.3-3.5). Both soybean parents produced lines that recovered or exceeded their yield. On the other hand, IA3023 and LG04-6000 were among the highest yielding soybean parents at 4958 kg/ha and 5054 kg/ha. No lines derived from these parents recovered the yield of the soybean parent, and these pedigrees also had lower average yields than progeny from Williams 82 and LN97-15076 (Table 3.6). One possible explanation for this phenomenon could be due to beneficial haplotype blocks that may be present in the higher yielding cultivars but absent in Williams 82 or LN97-15076. If this scenario were true, crosses with *G. soja* could decrease yields by disrupting these haplotype blocks.

Effects of Selection on Genomic Composition

The stringent selection imposed on the *G. max* x *G. soja* populations not only led to changes in phenotype, but also changes in genetic composition. While BC_0 populations would be expected to carry equal proportions of alleles from either parent, the average overall *G. soja* allele frequency in lines from BC_0 crosses only ranged from 16.6% to 26.9% (Table 3.9). This allele frequency is more similar to what is expected in a BC_1 population (25%). Selection was only performed based on a plant's phenotype, although correlations do exist between the total proportion of *G. soja* alleles and some important agronomic traits. Figures 3.2 and 3.3 display the relationship between yield and the amount of *G. soja* alleles in MS III and MS IV respectively. Due to the presence of significant environmental and genotype-by-environment effects, each test was analyzed separately since environments were not uniform across all three tests. Among the BC_0 crosses, only LG04-6000 x PI 549048 and Williams 82 x PI 479752 in MS III and MS IV contained enough lines to perform a regression analysis. In MS III, increased

contribution from the *G. soja* parent was associated with a decrease in yield for both crosses (-63.9 kg/ha in LG04-6000 x PI 549048 and -37.3 kg/ha in Williams 82; Fig. 3.2). However, this trend was not as strong among the MS IV lines (Fig. 3.3). Within the LG04-6000 x PI 549048 pedigree, yield decreased by 42.6 kg/ha (21.3 kg/ha less than MS III). Among the MS IV lines derived from Williams 82 x PI 479752, yield tended to decrease with increasing *G. soja* contribution but the response was not statistically significant (Fig. 3.3).

While a general negative trend was observed between yield and *G. soja* allele percentage, some lines can be observed with high values for both. For example, LG14-3365, and LG14-3200 were among the highest yielding Williams 82 x PI 479752 lines (3930, and 3833 kg/ha, respectively) while also containing 33%, and 31% *G. soja* alleles respectively. The combination of high yields and large genetic contributions from *G. soja* could make these lines desirable for introducing new diversity into soybean breeding. While the yields were not significantly different from Williams 82 (3930 kg/ha), the lines did differ for other agronomic traits including lodging and 100-seed weight (Table 3.5).

The amount of *G. soja* alleles was also related to 100-seed weight in the Williams 82 x PI 479752 pedigree (Fig. 3.4 and 3.5). Among lines in this pedigree, 100-seed weight decreased in MS III by 0.24 g for each percent increase in *G. soja* alleles (Fig. 3.4), and by 0.20 g in MS IV (Fig. 3.5). This was an expected relationship given the quantitative nature of seed weight in soybean and the large number of QTL associated with the trait. From the Williams 82 x PI 479752 mapping population in Chapter 2, ten significant seed weight QTL were detected on eight different chromosomes. Large seeded lines would be expected to contain *G. max* alleles at the majority of these loci along with any linked regions derived from the *G. max* parent, thereby making it difficult to maintain high levels of *G. soja* alleles with large seeds. Like yield however, the relationship between seed weight and the percentage of *G. soja* alleles was variable. Thirty-eight percent of the SNPs in LG14-3327 were derived from *G. soja*, yet the 100-seed weight (10.2 g) was equivalent to other lines with only 20% *G. soja* alleles. However, the same relationship was not observed in the LG04-6000 x PI 549048 pedigree. Among MS III lines from this pedigree, the regression coefficient was not statistically significant at α=0.05 (Fig. 3.4), and 100-seed weight slightly increased among MS IV lines by 0.04 g for each percent increase in *G. soja* alleles (Fig 3.5). A possible explanation for this pattern may lie in the limited range in seed weights within this pedigree. While Williams 82 x PI 479752 lines ranged from 14.3 g to 6.1 g,

the largest seeded line from LG04-6000 x PI 549048 was only 9.7 g. Without the largest seed weights represented, this may limit the ability to accurately detect the relationship between percent *G. soja* alleles and 100-seed weight.

The effect of selection can also be seen in the distribution of *G. soja* alleles in Williams 82 x PI 479752 lines (Fig. 3.1). Allele frequencies depicted in Fig. 3.1 are smoothed using a 20 SNP sliding window with a 10 SNP overlap between windows. Introgressions from PI 479752 were not randomly distributed across the genome, but were detected in regions of low and high frequency. Only a few regions contained *G. soja* alleles at frequencies near or above 50%, predominantly on chr 3, 4, 7, 8, 15, and 18. Areas of higher *G. soja* frequency most often occurred in regions without domestication-related QTL. This would indicate that these regions contain neutral or potentially positive alleles. With a few exceptions, all 32 regions containing DRT QTL were located at or adjacent to regions of low *G. soja* allele frequency (Fig. 3.1). On chr 12, *G. soja* alleles occurred at approximately 40% at *qSW-12*. This QTL was detected within the Williams 82 x PI 479752 population in Chapter 2 and explained 11% of the variation. The negative allele may have been allowed to persist among the entries in this study due to the relatively small effect of the QTL (-0.3 g per *G. soja* allele). Domestication-related QTL on chr 18 and *qLD-19-2* on chr 19 also did not appear to be under strong selection. These QTL only explained small portions of the variation in the mapping populations, and therefore were less correlated to undesirable phenotypes.

During the inbreeding process of the *G. soja*-derive lines, we selected strongly against shattering susceptibility. For an oligogenic trait such as shattering (Dong et al., 2014; Funatsuki et al., 2014), strong selection against a phenotype results in near-fixation at the locus. Figure 3.6 shows the allele frequencies surrounding *qSH-16*, a major QTL in Williams 82 x PI 479752. Due to its strong effect on shattering (53% variance explained), *qSH-16* was nearly fixed at 30 Mb for the *G. max* allele. Only two lines (LG14-3169 and LG14-3173) contained *G. soja* alleles at this locus, and were also the most shattering susceptible (48% and 35% respectively). The high frequency of shattering severely reduced the harvested yield of these lines which only averaged 943 kg/ha and 1728 kg/ha respectively. A total of 20 Williams 82 x PI 479752 lines contained *G. soja* introgressions within 500 kb of the *qSH-16* locus, and all displayed low levels of shattering (under 5%). Two lines with introgressions adjacent to *qSH-16*, LG14-3030 and LG14-3041, yielded 3793 kg/ha ad 3725 kg/ha, which was not significantly different from Williams 82 (3954

kg/ha). These lines could be useful for increasing diversity within the selective sweep at *qSH-16* without introducing undesirable traits.

Selection for 100-seed weight would have occurred only indirectly as agronomically good phenotypes would generally have larger seeds so recovering the 100-seed weight of the soybean parent was more difficult than reducing shattering. Lines from the Williams 82 x PI 479752 pedigree averaged 9 g/100 seeds, considerably less than the 16.1g 100-seed weight of Williams 82. Unlike shattering, 100-seed weight is controlled by many QTL with small to moderate effects. This caused weaker selection against negative *G. soja* alleles at the seed weight QTL. While *qSW-17-1* accounted for the largest phenotypic variance in the Williams 82 x PI 479752 mapping population in Chapter 2, the *G. soja* allele at that locus still persisted at approximately 25% (Fig. 3.7). Weak negative effects from the *G. soja* allele at one locus may be counteracted by positive effects of *G. max* alleles at other loci, thus inhibiting fixation. From Williams 82 x PI 479752, 31 lines had 100-seed weights of 10 g or more. Of these, eight contained *G. soja* introgressions within one or more of the major seed weight QTL, *qSW-12*, *qSW-17-1*, or *qSW-19* (Table 3.10). Two lines, LG14-3030 and LG14-3415, carried *G. soja* alleles at two of the three seed weight QTL while still maintaining 100-seed weights of 10 g. LG14-3030, LG14-3228, and LG14-3411 were not significantly different in yield from Williams 82, although all lines had smaller seeds than Williams 82

Comparisons to Similar Study

The plant material and objectives in this study were similar to a previous work published by Akpertey et al. (2014). Their study utilized 93 *G. soja*-derived lines from BC_1 and BC_2 populations to identify high-yielding lines and patterns of introgression from the wild parent. A few differences can be observed in the results from this current study and Akpertey et al. (2014). First, while three lines were identified that displayed yields above the higher yielding recurrent parent, the differences were not considered statistically significant. In this study, 22 lines were equivalent in yield to the soybean parent and two lines were higher yielding at $\alpha = 0.05$ (Tables 3.7 and 3.8). Three of these lines were either tested by Akpertey et al. (2014) (LG07-4231) or shared the same pedigree with lines in that study (LG14-3897 and LG14-3899). The larger number of lines evaluated in this study (416) likely contributed to the improved ability to identify *G. soja*-derived lines equivalent or above the *G. max* parent. The large majority of such

lines were derived from Williams 82 x PI 479752, a pedigree not utilized by Akpertey et al. (2014).

Second, Akpertey et al. (2014) found that the overall contribution from *G. soja* did not significantly deviate from what would be expected in backcross populations without selection. In BC_2 lines, the percentage of *G. soja* alleles ranged from 1 – 32%, however the average contribution was 13%. This indicated that phenotypic selection for agronomic traits did not significantly alter the overall proportion of wild soybean alleles from the expected 12.5%. In the present study, the majority of lines with genotyped parents were derived from BC_0 crosses (Table 3.9). Among these lines with 50% *G. soja* contribution, the percentage of detected *G. soja* SNPs averaged only 25%. The cross of Williams 82 x PI 479752 displayed the largest range with a minimum of 8.8% and maximum of 41.7%. All BC_0 crosses averaged under 50%, and the differences were significantly lower than what would be expected in a population without selection (Table 3.9). This result is more in line with our original hypothesis that selection for *G. max* phenotypes would lead to an increase in overall *G. max* alleles. A possible reason for the apparent discrepancy in results between the two studies may lie in the difference in backcross generations. While the average of *G. soja* alleles in BC_0 was significantly below the expected 50%, the BC_2 crosses were not significantly below 12.5% (7.6% in LN97-15076 x LG01-7770 and 12.3% in SS98-3403 (3) x PI 483461; Table 3.9). It is plausible that selection in BC generations has a decreasing effect on *G. soja* content as the number of back crosses increase. According to the study by Lam et al. (2010), only ~5% of the total soybean genome exhibits divergence from *G. soja* as measured by F_{ST}, a number similar to other species such as maize (Yamasaki et al., 2007). By decreasing the contribution of the wild parent, the probability of containing undesirable alleles within this 5% of the genome also decreases. While BC_2 and BC_0 crosses displayed different responses to phenotypic selection, the number of BC_2 lines with all parents genotyped (14) was considerably lower than BC_0 lines (206), which makes accurate comparisons between the two types of crosses difficult.

In the study by Akpertey et al. (2014), the authors were unable to detect SNPs for which *G. soja* alleles were consistently selected for or against. For important domestication traits, it would be expected to find linked loci that only contain the *G. max* allele. Patterns of selection were detected in the present study when examining Williams 82 x PI 479752 lines. Among these lines, *G. soja* allele frequencies under 20% were frequently detected, especially in regions

associated with domestication-related traits (Fig. 3.1). For example, low frequencies were detected on chr 16 in the region of a strong shattering QTL (Fig. 3.7) and on chr 13 in a region with QTL for lodging, main stem length, stem diameter, and growth habit. Because of the low number of markers (1536) used by Akpertey et al. (2014), these regions may have been distanced from any flanking markers such that linkage disequilibrium did not exist between the QTL and the markers. While regions of selection were identified in one BC_0 population from the current study (Williams 82 x PI 479752), patterns should also be examined in the additional pedigrees to identify universally selected regions.

Conclusion

This study represents one of only a few instances of recovering or exceeding the yield of a soybean parent using *G. soja*. Although the wild species represents a rich source of genetic diversity, regaining agronomic performance after crosses between the species is difficult. The high-yielding lines from this study can be used in a soybean breeding program to introduce potentially novel diversity from *G. soja* while avoiding the cost of screening large *G. max* x *G. soja* populations. These lines may also be useful in a breeding program to introduce new diversity within domestication-related selective sweeps (Zhou et al., 2015). While lines were typically fixed for *G. max* alleles at important DRT loci, introgressions of *G. soja* alleles did occur adjacent to the loci (Fig 3.2). Identifying lines with introgressions within selective sweeps may add additional value to using this material in a breeding program.

TABLES

Table 3.1. *G. max* x *G. soja* crosses evaluated in study. Percent contribution by pedigree for each parent given in parentheses.

Cross	BC[†]	*G. soja*	*G. max*			2015 Total[‡]	2016 Total[‡]
03JR309156 x LG07-2640	3	PI 483461 (6.25)	03JR09156 (50)	LN97-15076 (25)	Williams 82 (18.75)	31	25
03JR309156 x LG07-4727	2	PI 507807 (12.5)	03JR09156 (50)	Williams 82 (37.5)		43	32
06NB204846 x LG07-2526	3	PI 065549 (6.25)	06NB204846 (50)	IA3023 (25)	Williams 82 (18.75)	25	25
06NB204846 x LG07-4216	3	PI 549046 (6.25)	06NB204846 (50)	Williams 82 (43.75)		11	9
06NB204846 x LG07-4307	2	PI 479767 + PI 549046 (12.5)	06NB204846 (50)	Williams 82 (37.5)		26	26
IA2052 x LG01-7812	2	PI 483461 (12.5)	IA2052 (50)	Williams 82 (37.5)		4	4
IA3023 x LG01-7884	2	PI 549046 (12.5)	IA3023 (50)	Williams 82 (37.5)		1	1
IA3023 x PI 424068	0	PI 424068 (50)	IA3023 (50)			2	0
IA3023 x PI 468397A	0	PI 468397A (50)	IA3023 (50)			10	0
IA3023 x PI 522183A	0	PI 522183A (50)	IA3023 (50)			15	11
IA3023 x PI 549048	0	PI 549048 (50)	IA3023 (50)			11	5
LG04-5190 x PI 507788	0	PI 507788 (50)	LG04-5190 (50)			17	12
LG04-6000 x PI 507787	0	PI 507787 (50)	LG04-6000 (50)			22	20
LG04-6000 x PI 507774	0	PI 507774 (50)	LG04-6000 (50)			1	0
LG04-6000 x PI 549048	0	PI 549048 (50)	LG04-6000 (50)			47	35
LG97-7012 x PI 468399A	0	PI 468399A (50)	LG97-7012 (50)			4	0
LN97-15076 x LG01-7770	2	PI 483461 (12.5)	LN97-15076 (50)	Williams 82 (37.5)		7	7
Macon x PI 483461	0	PI 483461 (50)	Macon (50)			2	2
SS98-3403 (3) x PI 483461	2	PI 483461 (12.5)	SS98-3403 (87.5)			7	0
Williams 82 (2) x PI 479767	1	PI 479767 (25)	Williams 82 (75)			4	4
Williams 82 x LG01-7728	2	PI 479767 (12.5)	Williams 82 (87.5)			1	0
Williams 82 x LG01-7909	2	PI 549046 (12.5)	Williams 82 (87.5)			1	1
Williams 82 x PI 479752	0	PI 479752 (50)	Williams 82 (50)			124	81

[†] Number of backcrosses to *G. max*

[‡] Total number of evaluated lines from each cross

Table 3.2. *G. soja* accessions used as parents with the origin and maturity group listed.

Accession	Origin	MG[†]
PI 065549	Heilongjiang, China	II
PI 424068	Kangwon, South Korea	IV
PI 468397A	Shanxi, China	IV
PI 468399A	Shandong, China	IV
PI 479752	Jilin, China	I
PI 479767	Heilongjiang, China	I
PI 483461	Hebei, China	II
PI 507774	Primorye, Russia	II
PI 507787	Primorye, Russia	II
PI 507788	Primorye, Russia	III
PI 507807	Russia	0
PI 522183A	Heilongjiang, China	0
PI 549046	Shaanxi, China	IV
PI 549048	Beijing, China	III

[†] Maturity group

Table 3.3. Trait means for *G. soja*-derived lines and checks analyzed in MS II test.

Entry	Pedigree	*G. soja* Parent	Yield (kg/ha)	R1 (days)	R8 (days)	Hgt† (cm)	Ldg‡ (1 - 5)	GH§ (1 - 5)	SH❡ (%)	SdWt†† (g)	% *G. soja*
LD02-4485	Check		4449	19	105	99	2.2	1.0	0	13.6	
LG11-6210	Check		4951	25	114	107	2.7	1.0	0	-	
LG14-3818	F5 LG04-5190 x PI 507788	PI 507788	2919	-	105	102	2.7	1.0	0	9.3	-
LG14-3825	F5 LG04-5190 x PI 507788	PI 507788	2814	-	101	108	3.7	2.0	0	10.7	-
LG14-3826	F5 LG04-5190 x PI 507788	PI 507788	3587	22	112	101	3.3	1.1	0	9.3	-
LG14-3789	F5 LG04-6000 x PI 507787	PI 507787	2978	-	111	114	4.2	2.0	0	9.8	27
LG11-9177	F5 Williams 82 (2) x PI 479767	PI 479767	3605	27	110	113	2.6	1.5	0	11.6	-
LG14-5005	F6 03JR309156 x LG07-2640	PI 483461	4523	25	110	103	2.1	1.0	0	13.4	-
LG14-5006	F6 03JR309156 x LG07-2640	PI 483461	4637	22	110	111	2.3	1.0	0	14.8	-
LG14-5007	F6 03JR309156 x LG07-2640	PI 483461	4650	25	111	104	1.9	1.0	0	13.7	-
LG14-5019	F6 03JR309156 x LG07-2640	PI 483461	4492	24	110	110	2.5	1.5	0	12.6	-
LG14-5023	F6 03JR309156 x LG07-2640	PI 483461	4557	27	112	113	2.6	1.4	0	14.0	-
LG14-6052	F6 03JR309156 x LG07-2640	PI 483461	4922	22	112	118	2.5	1.3	0	14.9	-
LG14-6061	F6 03JR309156 x LG07-2640	PI 483461	4602	23	110	105	2.4	1.3	0	13.9	-
LG14-6064	F6 03JR309156 x LG07-2640	PI 483461	4741	22	111	111	2.5	1.1	0	14.0	-
LG14-5086	F6 06NB204846 x LG07-2526	PI 065549	4424	24	108	114	2.8	1.8	0	13.3	-
LG14-5087	F6 06NB204846 x LG07-2526	PI 065549	4583	24	108	113	3.4	2.0	0	13.6	-
LG14-5092	F6 06NB204846 x LG07-2526	PI 065549	4091	23	106	102	2.3	1.5	0	14.0	-
LG14-5094	F6 06NB204846 x LG07-2526	PI 065549	4574	21	106	102	2.8	1.6	0	13.8	-
LG14-5098	F6 06NB204846 x LG07-2526	PI 065549	4125	22	105	108	2.8	1.6	0	15.1	-
LG14-5100	F6 06NB204846 x LG07-2526	PI 065549	4789	25	111	103	2.0	1.0	0	16.0	-
LG14-5102	F6 06NB204846 x LG07-2526	PI 065549	4319	26	111	105	2.1	1.0	0	16.3	-
LG14-5103	F6 06NB204846 x LG07-2526	PI 065549	4429	27	109	104	2.3	1.0	0	17.2	-
LG14-5107	F6 06NB204846 x LG07-2526	PI 065549	4116	23	110	94	2.1	1.0	0	17.2	-
LG14-5108	F6 06NB204846 x LG07-2526	PI 065549	4505	24	109	102	2.1	1.0	0	19.1	-
LSD (p < 0.05)			**318**	**2**	**1**	**6**	**0.4**	**0.2**	**3**	**0.6**	

Table 3.3. (cont.)

Entry	Pedigree	*G. soja* Parent	Yield (kg/ha)	R1 (days)	R8 (days)	Hgt† (cm)	Ldg‡ (1 - 5)	GH§ (1 - 5)	SH¶ (%)	SdWt†† (g)	% *G. soja*
LG14-5110	F6 06NB204846 x LG07-2526	PI 065549	4631	24	112	108	2.7	1.2	0	15.6	-
LG14-5116	F6 06NB204846 x LG07-2526	PI 065549	3828	27	107	104	2.6	1.1	0	18.3	-
LG14-5117	F6 06NB204846 x LG07-2526	PI 065549	4472	24	106	103	2.5	1.6	0	16.7	-
LG14-5119	F6 06NB204846 x LG07-2526	PI 065549	4059	26	108	106	2.9	1.3	0	16.8	-
LG14-5123	F6 06NB204846 x LG07-2526	PI 065549	4540	25	107	106	2.6	1.4	0	17.5	-
LG14-5124	F6 06NB204846 x LG07-2526	PI 065549	4294	30	109	111	2.9	1.3	0	18.0	-
LG14-5127	F6 06NB204846 x LG07-2526	PI 065549	4081	24	106	102	2.9	1.4	0	16.5	-
LG14-5128	F6 06NB204846 x LG07-2526	PI 065549	3949	25	110	107	2.7	1.3	0	16.5	-
LG14-5129	F6 06NB204846 x LG07-2526	PI 065549	3903	26	107	102	2.7	1.2	0	17.2	-
LG14-5130	F6 06NB204846 x LG07-2526	PI 065549	4680	22	107	109	2.6	1.3	0	16.3	-
LG14-5134	F6 06NB204846 x LG07-2526	PI 065549	4364	23	108	100	2.7	1.3	0	18.2	-
LG14-5159	F6 06NB204846 x LG07-4216	PI 549046	3900	-	108	101	2.7	1.0	0	13.0	-
LG14-5173	F6 06NB204846 x LG07-4216	PI 549046	4084	-	110	115	3.0	1.5	0	13.4	-
LG14-5178	F6 06NB204846 x LG07-4307	PI 479767 + PI 549046	4760	20	109	103	2.3	1.0	0	14.3	-
LG14-5192	F6 06NB204846 x LG07-4307	PI 479767 + PI 549046	4441	24	107	108	2.6	1.3	0	13.6	-
LG14-5193	F6 06NB204846 x LG07-4307	PI 479767 + PI 549046	4304	23	112	106	2.8	1.0	0	12.1	-
LG14-5194	F6 06NB204846 x LG07-4307	PI 479767 + PI 549046	4494	25	108	103	2.1	1.2	0	12.1	-
LG14-5205	F6 06NB204846 x LG07-4307	PI 479767 + PI 549046	4366	23	111	116	2.7	1.3	0	13.7	-
LG14-5208	F6 06NB204846 x LG07-4307	PI 479767 + PI 549046	4471	22	108	102	2.1	1.1	0	16.5	-
LG14-5209	F6 06NB204846 x LG07-4307	PI 479767 + PI 549046	4158	24	111	118	2.6	1.1	0	13.2	-
LG14-5210	F6 06NB204846 x LG07-4307	PI 479767 + PI 549046	4604	24	112	114	2.7	1.2	0	15.2	-
LG14-6260	F6 06NB204846 x LG07-4307	PI 479767 + PI 549046	4642	22	110	113	2.4	1.4	0	14.4	-
LSD (p < 0.05)			**318**	**2**	**1**	**6**	**0.4**	**0.2**	**3**	**0.6**	

Table 3.3. (cont.)

Entry	Pedigree	*G. soja* Parent	Yield (kg/ha)	R1 (days)	R8 (days)	Hgt† (cm)	Ldg‡ (1 - 5)	GH§ (1 - 5)	SH‖ (%)	SdWt†† (g)	% *G. soja*
LG14-6283	F6 06NB204846 x LG07-4307	PI 479767 + PI 549046	4393	21	108	106	2.1	1.1	0	13.4	-
LG14-6286	F6 06NB204846 x LG07-4307	PI 479767 + PI 549046	4414	21	110	112	2.3	1.1	0	14.3	-
LG14-6290	F6 06NB204846 x LG07-4307	PI 479767 + PI 549046	4341	22	108	103	2.2	1.0	0	14.2	-
LG14-6296	F6 06NB204846 x LG07-4307	PI 479767 + PI 549046	4107	22	111	105	2.8	1.1	0	11.0	-
LG14-6297	F6 06NB204846 x LG07-4307	PI 479767 + PI 549046	4126	23	109	107	2.6	1.3	0	13.7	-
LG14-6301	F6 06NB204846 x LG07-4307	PI 479767 + PI 549046	4295	24	109	111	2.5	1.1	0	14.6	-
LG14-2806	F6 IA3023 x PI424068	PI 424068	3036	-	115	101	3.6	1.3	0	7.1	-
LG14-2699	F6 IA3023 x PI468397A	PI 468397A	2938	42	113	114	4.5	2.0	0	5.5	15
LG14-2716	F6 IA3023 x PI468397A	PI 468397A	2228	23	117	116	4.5	1.5	10	6.0	14
LG14-2721	F6 IA3023 x PI468397A	PI 468397A	3062	36	105	104	3.5	2.0	5	6.7	12
LG14-2756	F6 IA3023 x PI468397A	PI 468397A	2846	45	118	122	4.5	2.5	0	5.1	16
LG14-2582	F6 IA3023 x PI522183A	PI 522183A	2064	21	108	96	3.3	1.3	40	6.9	24
LG14-2588	F6 IA3023 x PI522183A	PI 522183A	1389	-	110	89	3.3	1.3	30	6.8	24
LG14-2620	F6 IA3023 x PI522183A	PI 522183A	1980	27	108	94	3.8	1.8	35	8.0	21
LG14-2792	F6 IA3023 x PI549048	PI 549048	2784	27	108	101	3.9	1.8	0	8.1	17
LG14-2795	F6 IA3023 x PI549048	PI 549048	2912	27	109	98	3.6	1.6	3	7.5	19
LG14-2797	F6 IA3023 x PI549048	PI 549048	2983	31	109	96	3.8	1.4	0	8.5	-
LG14-3478	F6 LG04-6000 x PI549048	PI 549048	3198	41	111	93	3.7	1.1	0	7.6	-
LG14-3548	F6 LG04-6000 x PI549048	PI 549048	2995	24	-	-	3.3	1.0	0	7.2	28
LG14-3633	F6 LG04-6000 x PI549048	PI 549048	3103	20	113	110	4.2	2.0	0	6.9	21
LG14-2834	F6 LG97-7012 x PI468399A	PI 468399A	1548	-	95	93	4.0	2.0	5	8.1	-
LG14-2848	F6 LG97-7012 x PI468399A	PI 468399A	2621	-	112	100	4.2	1.5	0	7.2	-
LSD (p < 0.05)			**318**	**2**	**1**	**6**	**0.4**	**0.2**	**3**	**0.6**	

Table 3.3. (cont.)

Entry	Pedigree	*G. soja* Parent	Yield (kg/ha)	R1 (days)	R8 (days)	Hgt† (cm)	Ldg‡ (1 - 5)	GH§ (1 - 5)	SH⁋ (%)	SdWt†† (g)	% *G. soja*
LG14-2855	F6 LG97-7012 x PI468399A	PI 468399A	3082	-	105	95	3.7	1.5	0	10.5	-
LG14-3886	F7 IA2052 x LG01-7812	PI 483461	4077	21	108	111	3.0	1.1	0	13.9	-
LG14-3887	F7 IA2052 x LG01-7812	PI 483461	4524	21	110	115	2.9	1.1	0	12.7	-
LG14-3890	F7 IA2052 x LG01-7812	PI 483461	4324	21	110	109	3.6	1.3	0	12.9	-
LG14-3892	F7 IA2052 x LG01-7812	PI 483461	4263	20	108	104	2.3	1.0	0	17.0	-
LG14-3067	F7 Williams 82 x PI 479752	PI 479752	3216	18	108	84	3.5	1.1	0	7.7	33
LG14-3073	F7 Williams 82 x PI 479752	PI 479752	3740	27	108	114	3.5	1.8	2	12.3	-
LG14-3099	F7 Williams 82 x PI 479752	PI 479752	2578	27	105	83	2.6	1.0	16	8.2	30
LG14-3199	F7 Williams 82 x PI 479752	PI 479752	3902	-	121	108	4.5	1.0	0	6.6	31
LG14-3276	F7 Williams 82 x PI 479752	PI 479752	2959	19	110	98	3.2	1.3	0	7.4	31
LG14-3327	F7 Williams 82 x PI 479752	PI 479752	2326	-	97	93	3.5	1.5	0	10.2	37
LG14-3328	F7 Williams 82 x PI 479752	PI 479752	2605	23	104	93	3.0	1.2	0	9.4	36
LG14-3340	F7 Williams 82 x PI 479752	PI 479752	2434	-	98	85	3.5	1.3	0	7.0	35
LG14-3372	F7 Williams 82 x PI 479752	PI 479752	3201	22	110	101	3.3	1.4	0	7.4	20
LG14-3431	F7 Williams 82 x PI 479752	PI 479752	1955	-	-	94	4.2	1.5	5	10.4	27
LG14-3451	F7 Williams 82 x PI 479752	PI 479752	3621	23	-	101	4.5	1.5	0	9.1	27
LG14-3458	F7 Williams 82 x PI 479752	PI 479752	739	-	106	58	1.5	1.0	0	12.2	14
LG14-3463	F7 Williams 82 x PI 479752	PI 479752	4197	22	108	98	2.1	1.2	0	13.1	15
LG14-3464	F7 Williams 82 x PI 479752	PI 479752	4127	22	111	101	1.9	1.0	0	13.2	15
LSD (p < 0.05)			**318**	**2**	**1**	**6**	**0.4**	**0.2**	**3**	**0.6**	

† Plant height

‡ Lodging

§ Growth Habit

⁋ Shattering

†† 100-Seed Weight

Table 3.4. Trait means for *G. soja*-derived lines and checks analyzed in MS III test.

Entry	Pedigree	*G. soja* Parent	Yield (kg/ha)	R1 (days)	R8 (days)	Hgt† (cm)	Ldg‡ (1 - 5)	GH§ (1 - 5)	SH‖ (%)	SdWt†† (g)	% *G. soja*
IA3023	Check		4958	27	113	103	2.0	1.0	0	16.1	
LD08-8622	Check		5184	24	112	108	2.6	1.1	0	15.8	
LG11-6210	Check		5271	28	116	110	2.5	1.0	0	-	
SS98-3403	Check		4202	29	115	109	1.7	1.0	0	17.1	
Williams 82	Check		3954	28	118	128	2.9	1.2	0	16.1	
LG14-3813	F5 LG04-5190 x PI 507788	PI 507788	3165	25	118	110	4.7	1.5	0	9.8	-
LG14-3814	F5 LG04-5190 x PI 507788	PI 507788	3288	-	125	109	4.1	1.3	0	7.2	-
LG14-3815	F5 LG04-5190 x PI 507788	PI 507788	3141	25	115	117	4.1	1.4	5	8.9	-
LG14-3824	F5 LG04-5190 x PI 507788	PI 507788	3332	29	113	108	3.6	1.6	2	9.4	-
LG14-3827	F5 LG04-5190 x PI 507788	PI 507788	3563	24	118	112	3.9	1.6	1	10.6	-
LG14-3829	F5 LG04-5190 x PI 507788	PI 507788	3309	24	115	112	2.9	1.0	2	9.2	-
LG14-3832	F5 LG04-5190 x PI 507788	PI 507788	2905	28	116	107	3.4	1.0	33	9.3	-
LG14-3834	F5 LG04-5190 x PI 507788	PI 507788	2953	29	122	120	3.9	1.3	15	8.7	-
LG14-3835	F5 LG04-5190 x PI 507788	PI 507788	2472	29	116	97	3.6	1.3	45	7.2	-
LG14-3841	F5 LG04-5190 x PI 507788	PI 507788	2930	24	115	103	4.0	1.4	1	6.9	-
LG14-3844	F5 LG04-5190 x PI 507788	PI 507788	2923	23	117	93	4.2	1.0	0	5.8	-
LG14-3848	F5 LG04-5190 x PI 507788	PI 507788	3132	24	117	109	4.1	1.5	0	5.5	-
LG14-3721	F5 LG04-6000 x PI 507787	PI 507787	2938	28	115	103	3.6	1.6	7	6.2	27
LG14-3724	F5 LG04-6000 x PI 507787	PI 507787	3058	29	114	118	3.8	1.6	0	7.0	27
LG14-3725	F5 LG04-6000 x PI 507787	PI 507787	3338	30	115	124	4.2	2.0	4	8.5	23
LG14-3737	F5 LG04-6000 x PI 507787	PI 507787	3872	29	118	111	3.2	1.3	0	11.0	17
LG14-3738	F5 LG04-6000 x PI 507787	PI 507787	3384	29	118	118	3.5	1.3	0	10.5	-
LG14-3759	F5 LG04-6000 x PI 507787	PI 507787	3653	30	112	114	3.9	1.9	0	10.5	27
LG14-3762	F5 LG04-6000 x PI 507787	PI 507787	3562	30	113	110	4.1	2.0	1	8.8	26
LG14-3767	F5 LG04-6000 x PI 507787	PI 507787	3914	25	115	105	4.1	1.6	9	11.5	22
LG14-3792	F5 LG04-6000 x PI 507787	PI 507787	3748	26	114	122	3.6	2.0	3	10.7	23
LSD ($p < 0.05$)			**269**	**1**	**1**	**7**	**0.3**	**0.2**	**4**	**0.7**	

Table 3.4. (cont.)

Entry	Pedigree	*G. soja* Parent	Yield (kg/ha)	R1 (days)	R8 (days)	Hgt† (cm)	Ldg‡ (1 - 5)	GH§ (1 - 5)	SH‖ (%)	SdWt†† (g)	% *G. soja*
LG14-3793	F5 LG04-6000 x PI 507787	PI 507787	3489	24	114	124	3.8	2.1	1	12.0	22
LG14-3801	F5 LG04-6000 x PI 507787	PI 507787	3181	28	120	118	3.4	1.2	1	9.7	-
LG14-3806	F5 LG04-6000 x PI 507787	PI 507787	2932	25	116	92	3.3	1.1	17	8.0	26
LG14-3470	F5 LG04-6000 x PI507774	PI 507774	2714	25	124	109	2.8	1.0	0	7.4	-
LG11-9161	F5 Williams 82 (2) x PI 479767	PI 479767	3934	25	113	123	3.0	1.3	0	14.2	-
LG11-9166	F5 Williams 82 (2) x PI 479767	PI 479767	3893	30	115	118	3.0	1.1	0	12.5	-
LG11-9189	F5 Williams 82 (2) x PI 479767	PI 479767	4138	28	115	120	3.0	1.3	0	13.5	-
LG07-4231	F5 Williams 82 x LG01-7909	PI 549046	4143	29	117	134	3.2	1.3	0	14.1	-
LG12-9591	F5 Williams 82 x PI 479752	PI 479752	4292	25	117	111	2.5	1.0	0	12.5	15
LG12-9619	F5 Williams 82 x PI 479752	PI 479752	4081	24	117	117	2.5	1.0	0	12.3	-
LG14-5001	F6 03JR309156 x LG07-2640	PI 483461	4617	26	111	116	2.9	1.5	0	14.7	-
LG14-5009	F6 03JR309156 x LG07-2640	PI 483461	4732	26	115	110	2.8	1.0	0	15.6	-
LG14-5015	F6 03JR309156 x LG07-2640	PI 483461	4393	28	117	115	3.4	1.3	0	13.7	-
LG14-5017	F6 03JR309156 x LG07-2640	PI 483461	4578	29	112	111	3.0	1.1	0	12.5	-
LG14-5021	F6 03JR309156 x LG07-2640	PI 483461	4426	28	113	106	3.1	1.0	0	12.1	-
LG14-5022	F6 03JR309156 x LG07-2640	PI 483461	4682	30	115	118	2.9	1.0	0	13.5	-
LG14-5024	F6 03JR309156 x LG07-2640	PI 483461	4332	31	116	114	3.2	1.3	0	13.3	-
LG14-6051	F6 03JR309156 x LG07-2640	PI 483461	4778	27	116	117	2.9	1.2	0	14.6	-
LG14-6055	F6 03JR309156 x LG07-2640	PI 483461	4720	-	112	111	3.0	1.0	0	15.1	-
LG14-6056	F6 03JR309156 x LG07-2640	PI 483461	4979	29	118	115	2.5	1.0	0	13.8	-
LG14-6058	F6 03JR309156 x LG07-2640	PI 483461	4895	25	115	115	2.7	1.0	0	12.8	-
LG14-6059	F6 03JR309156 x LG07-2640	PI 483461	5126	27	116	117	3.0	1.1	0	14.8	-
LG14-6067	F6 03JR309156 x LG07-2640	PI 483461	4943	28	119	111	2.8	1.0	0	14.4	-
LG14-6082	F6 03JR309156 x LG07-2640	PI 483461	4833	25	119	108	3.2	1.0	0	15.5	-
LG14-6085	F6 03JR309156 x LG07-2640	PI 483461	5069	25	117	109	2.6	1.1	0	16.2	-
LSD (p < 0.05)			**269**	**1**	**1**	**7**	**0.3**	**0.2**	**4**	**0.7**	

Table 3.4. (cont.)

Entry	Pedigree	*G. soja* Parent	Yield (kg/ha)	R1 (days)	R8 (days)	Hgt† (cm)	Ldg‡ (1 - 5)	GH§ (1 - 5)	SH⁋ (%)	SdWt†† (g)	% *G. soja*
LG14-6088	F6 03JR309156 x LG07-2640	PI 483461	4860	28	114	123	3.1	1.5	0	16.1	-
LG14-6089	F6 03JR309156 x LG07-2640	PI 483461	5121	29	120	123	2.9	1.3	0	16.3	-
LG14-6090	F6 03JR309156 x LG07-2640	PI 483461	5059	26	117	107	1.8	1.0	0	17.0	-
LG14-6091	F6 03JR309156 x LG07-2640	PI 483461	5209	30	118	121	2.5	1.1	0	17.5	-
LG14-6096	F6 03JR309156 x LG07-2640	PI 483461	5275	29	116	115	2.4	1.0	0	17.3	-
LG14-6099	F6 03JR309156 x LG07-2640	PI 483461	5295	30	116	111	2.4	1.0	0	15.6	-
LG14-6101	F6 03JR309156 x LG07-2640	PI 483461	4774	28	113	110	3.0	1.5	0	14.9	-
LG14-6102	F6 03JR309156 x LG07-4727	PI 507807	4303	-	115	103	2.3	1.0	0	16.5	-
LG14-6103	F6 03JR309156 x LG07-4727	PI 507807	4232	25	113	113	2.4	1.1	0	18.3	-
LG14-6106	F6 03JR309156 x LG07-4727	PI 507807	4780	24	113	104	1.9	1.0	0	17.3	-
LG14-6109	F6 03JR309156 x LG07-4727	PI 507807	4653	24	117	111	3.1	1.0	0	17.8	-
LG14-6110	F6 03JR309156 x LG07-4727	PI 507807	4637	25	116	117	3.0	1.0	0	16.9	-
LG14-6111	F6 03JR309156 x LG07-4727	PI 507807	4444	-	116	117	2.3	1.0	0	17.6	-
LG14-6113	F6 03JR309156 x LG07-4727	PI 507807	4918	27	113	103	2.3	1.1	0	16.4	-
LG14-6118	F6 03JR309156 x LG07-4727	PI 507807	4534	23	114	112	2.5	1.2	0	16.3	-
LG14-6119	F6 03JR309156 x LG07-4727	PI 507807	4553	-	111	117	2.3	1.0	0	17.4	-
LG14-6120	F6 03JR309156 x LG07-4727	PI 507807	4305	25	112	109	2.5	1.0	0	16.6	-
LG14-6122	F6 03JR309156 x LG07-4727	PI 507807	4705	24	112	104	2.0	1.1	0	17.4	-
LG14-6138	F6 03JR309156 x LG07-4727	PI 507807	4541	26	113	110	1.9	1.0	0	15.9	-
LG14-6139	F6 03JR309156 x LG07-4727	PI 507807	4439	26	115	111	2.4	1.0	0	16.7	-
LG14-6140	F6 03JR309156 x LG07-4727	PI 507807	4743	25	116	108	1.9	1.0	0	15.7	-
LG14-6142	F6 03JR309156 x LG07-4727	PI 507807	4584	26	117	113	2.3	1.0	0	16.1	-
LG14-6148	F6 03JR309156 x LG07-4727	PI 507807	4718	29	117	119	3.0	1.0	0	14.5	-
LG14-6150	F6 03JR309156 x LG07-4727	PI 507807	4498	25	114	116	2.7	1.0	0	13.5	-
LG14-6151	F6 03JR309156 x LG07-4727	PI 507807	4357	29	115	117	3.0	1.2	0	14.1	-
LG14-6152	F6 03JR309156 x LG07-4727	PI 507807	4776	29	118	118	2.9	1.0	0	16.4	-
LSD ($p < 0.05$)			**269**	**1**	**1**	**7**	**0.3**	**0.2**	**4**	**0.7**	

Table 3.4. (cont.)

Entry	Pedigree	*G. soja* Parent	Yield (kg/ha)	R1 (days)	R8 (days)	Hgt† (cm)	Ldg‡ (1 - 5)	GH§ (1 - 5)	SH⁋ (%)	SdWt†† (g)	% *G. soja*
LG14-6153	F6 03JR309156 x LG07-4727	PI 507807	4692	26	115	115	2.9	1.1	0	16.6	-
LG14-6155	F6 03JR309156 x LG07-4727	PI 507807	4379	28	117	117	3.1	1.1	0	14.9	-
LG14-6159	F6 03JR309156 x LG07-4727	PI 507807	4687	30	117	117	3.1	1.0	0	15.9	-
LG14-6160	F6 03JR309156 x LG07-4727	PI 507807	4265	26	119	135	3.4	1.0	0	16.6	-
LG14-6161	F6 03JR309156 x LG07-4727	PI 507807	4281	26	119	110	2.7	1.0	0	16.5	-
LG14-7926	F6 03JR309156 x LG07-4727	PI 507807	4590	28	117	121	3.3	1.1	0	16.5	-
LG14-7928	F6 03JR309156 x LG07-4727	PI 507807	4759	29	119	126	2.8	1.0	0	16.8	-
LG14-7932	F6 03JR309156 x LG07-4727	PI 507807	4585	24	117	110	2.1	1.0	0	18.8	-
LG14-7935	F6 03JR309156 x LG07-4727	PI 507807	4650	24	118	117	2.9	1.0	0	16.2	-
LG14-7941	F6 03JR309156 x LG07-4727	PI 507807	4330	24	116	115	2.8	1.0	0	16.0	-
LG14-7942	F6 03JR309156 x LG07-4727	PI 507807	4422	23	116	125	3.3	1.4	0	16.5	-
LG14-7945	F6 03JR309156 x LG07-4727	PI 507807	4812	25	117	116	3.0	1.1	0	16.0	-
LG14-5109	F6 06NB204846 x LG07-2526	PI 065549	4787	26	113	118	2.5	1.1	0	18.4	-
LG14-5112	F6 06NB204846 x LG07-2526	PI 065549	4721	24	112	102	2.7	1.0	0	17.0	-
LG14-5121	F6 06NB204846 x LG07-2526	PI 065549	4229	24	110	110	3.0	1.0	0	18.2	-
LG14-5143	F6 06NB204846 x LG07-2526	PI 065549	4663	26	111	108	3.0	1.5	0	15.8	-
LG14-5148	F6 06NB204846 x LG07-4216	PI 549046	3742	23	111	123	3.3	1.6	0	15.3	-
LG14-5149	F6 06NB204846 x LG07-4216	PI 549046	3516	24	111	120	3.7	1.8	0	15.4	-
LG14-5150	F6 06NB204846 x LG07-4216	PI 549046	3937	22	113	134	3.5	1.7	0	16.9	-
LG14-5152	F6 06NB204846 x LG07-4216	PI 549046	3698	23	112	138	3.5	1.9	0	15.7	-
LG14-5156	F6 06NB204846 x LG07-4216	PI 549046	3665	23	112	128	3.5	1.9	0	15.9	-
LG14-5169	F6 06NB204846 x LG07-4216	PI 549046	3909	28	112	132	3.7	2.3	0	13.6	-
LG14-5171	F6 06NB204846 x LG07-4216	PI 549046	4114	28	112	132	3.7	2.3	0	13.7	-
LG14-5172	F6 06NB204846 x LG07-4216	PI 549046	3999	26	112	137	3.7	2.3	0	13.2	-
LG14-5174	F6 06NB204846 x LG07-4216	PI 549046	4061	26	111	124	3.4	2.0	0	12.7	-
LSD (p < 0.05)			**269**	**1**	**1**	**7**	**0.3**	**0.2**	**4**	**0.7**	

Table 3.4. (cont.)

Entry	Pedigree	*G. soja* Parent	Yield (kg/ha)	R1 (days)	R8 (days)	Hgt† (cm)	Ldg‡ (1 - 5)	GH§ (1 - 5)	SH‖ (%)	SdWt†† (g)	% *G. soja*
LG14-6246	F6 06NB204846 x LG07-4307	PI 479767+ PI 549046	4720	26	114	117	3.0	1.4	0	15.8	-
LG14-6249	F6 06NB204846 x LG07-4307	PI 479767+ PI 549046	4729	23	115	123	3.0	1.1	0	13.9	-
LG14-6251	F6 06NB204846 x LG07-4307	PI 479767+ PI 549046	4824	25	112	112	3.1	1.5	0	15.4	-
LG14-6257	F6 06NB204846 x LG07-4307	PI 479767+ PI 549046	5082	25	116	119	3.3	1.2	0	15.6	-
LG14-6262	F6 06NB204846 x LG07-4307	PI 479767+ PI 549046	4553	26	118	116	3.2	1.1	0	17.1	-
LG14-6266	F6 06NB204846 x LG07-4307	PI 479767+ PI 549046	4579	26	112	111	3.1	1.3	0	11.9	-
LG14-6270	F6 06NB204846 x LG07-4307	PI 479767+ PI 549046	5087	27	114	113	2.4	1.1	0	17.4	-
LG14-6271	F6 06NB204846 x LG07-4307	PI 479767+ PI 549046	5192	25	112	112	2.7	1.2	0	18.7	-
LG14-6292	F6 06NB204846 x LG07-4307	PI 479767+ PI 549046	4456	23	111	111	2.9	1.0	0	12.5	-
LG14-6305	F6 06NB204846 x LG07-4307	PI 479767+ PI 549046	4570	26	112	109	2.5	1.0	0	13.9	-
LG14-6307	F6 06NB204846 x LG07-4307	PI 479767+ PI 549046	4375	26	113	115	2.6	1.0	0	12.0	-
LG08-3465	F6 IA3023 x LG01-7884	PI 549046	4587	29	119	118	2.8	1.1	0	16.6	-
LG14-2809	F6 IA3023 x PI424068	PI 424068	3046	-	115	109	3.6	1.0	0	7.5	-
LG14-2704	F6 IA3023 x PI468397A	PI 468397A	3516	48	117	137	4.1	2.0	0	6.8	13
LG14-2706	F6 IA3023 x PI468397A	PI 468397A	3088	45	112	149	4.1	1.8	0	6.6	15
LG14-2710	F6 IA3023 x PI468397A	PI 468397A	2917	26	112	116	3.2	1.0	0	6.8	14
LG14-2773	F6 IA3023 x PI468397A	PI 468397A	2970	52	115	109	4.1	1.8	0	5.9	23
LG14-2776	F6 IA3023 x PI468397A	PI 468397A	2769	49	115	97	3.9	1.5	0	5.6	22
LG14-2777	F6 IA3023 x PI468397A	PI 468397A	2686	50	118	95	4.1	1.3	0	5.7	22
LG14-2573	F6 IA3023 x PI522183A	PI 522183A	2922	38	118	65	3.1	1.1	10	8.9	27
LG14-2585	F6 IA3023 x PI522183A	PI 522183A	1023	41	116	112	3.8	2.0	38	6.4	26
LSD ($p < 0.05$)			**269**	**1**	**1**	**7**	**0.3**	**0.2**	**4**	**0.7**	

Table 3.4. (cont.)

Entry	Pedigree	*G. soja* Parent	Yield (kg/ha)	R1 (days)	R8 (days)	Hgt† (cm)	Ldg‡ (1 - 5)	GH§ (1 - 5)	SH‖ (%)	SdWt†† (g)	% *G. soja*
LG14-2602	F6 IA3023 x PI522183A	PI 522183A	2753	29	114	110	4.1	2.0	8	7.5	20
LG14-2605	F6 IA3023 x PI522183A	PI 522183A	3290	29	116	112	3.5	1.9	0	9.4	17
LG14-2607	F6 IA3023 x PI522183A	PI 522183A	2973	28	117	116	3.6	1.4	0	9.1	19
LG14-2613	F6 IA3023 x PI522183A	PI 522183A	3163	30	120	120	4.6	2.0	0	7.3	23
LG14-2615	F6 IA3023 x PI522183A	PI 522183A	2354	33	111	112	3.6	2.0	20	10.4	-
LG14-2618	F6 IA3023 x PI522183A	PI 522183A	2015	30	110	103	4.1	1.6	34	8.4	22
LG14-2655	F6 IA3023 x PI522183A	PI 522183A	3017	48	122	108	4.0	1.8	0	7.4	17
LG14-2794	F6 IA3023 x PI549048	PI 549048	3196	29	112	97	3.9	1.8	1	8.3	21
LG14-3484	F6 LG04-6000 x PI549048	PI 549048	3428	44	117	102	2.1	1.0	8	7.0	23
LG14-3488	F6 LG04-6000 x PI549048	PI 549048	3544	45	116	103	2.6	1.0	0	6.6	25
LG14-3492	F6 LG04-6000 x PI549048	PI 549048	3214	44	117	101	2.8	1.1	23	6.1	26
LG14-3533	F6 LG04-6000 x PI549048	PI 549048	2745	-	119	109	3.4	1.3	5	6.9	29
LG14-3536	F6 LG04-6000 x PI549048	PI 549048	3290	44	116	97	3.5	1.1	11	6.8	26
LG14-3551	F6 LG04-6000 x PI549048	PI 549048	3235	42	114	111	3.6	1.8	0	6.8	26
LG14-3553	F6 LG04-6000 x PI549048	PI 549048	3361	43	113	112	3.7	1.3	13	7.7	21
LG14-3557	F6 LG04-6000 x PI549048	PI 549048	3543	26	115	101	2.3	1.0	0	7.7	22
LG14-3559	F6 LG04-6000 x PI549048	PI 549048	3614	45	115	105	2.4	1.1	0	7.7	24
LG14-3572	F6 LG04-6000 x PI549048	PI 549048	3570	45	114	110	2.5	1.6	1	7.6	26
LG14-3578	F6 LG04-6000 x PI549048	PI 549048	3424	46	115	110	2.2	1.0	0	7.8	27
LG14-3594	F6 LG04-6000 x PI549048	PI 549048	2845	46	115	110	3.4	1.1	23	7.4	25
LG14-3606	F6 LG04-6000 x PI549048	PI 549048	2862	27	114	106	4.1	1.6	6	7.1	28
LG14-3623	F6 LG04-6000 x PI549048	PI 549048	3501	27	114	112	4.2	1.3	3	6.6	25
LG14-3647	F6 LG04-6000 x PI549048	PI 549048	2957	28	123	118	4.0	1.3	0	8.8	33
LG14-3651	F6 LG04-6000 x PI549048	PI 549048	2909	25	123	118	4.0	1.0	0	8.8	-
LG14-3655	F6 LG04-6000 x PI549048	PI 549048	2864	25	120	114	3.8	1.3	0	6.7	30
LG14-3712	F6 LG04-6000 x PI549048	PI 549048	3380	44	114	76	3.3	1.3	0	7.9	23
LSD (p < 0.05)			**269**	**1**	**1**	**7**	**0.3**	**0.2**	**4**	**0.7**	

Table 3.4. (cont.)

Entry	Pedigree	*G. soja* Parent	Yield (kg/ha)	R1 (days)	R8 (days)	Hgt† (cm)	Ldg‡ (1 - 5)	GH§ (1 - 5)	SH‖ (%)	SdWt†† (g)	% *G. soja*
LG11-9033	F6 Macon x PI 483461	PI 483461	3920	44	117	129	3.8	1.2	0	10.0	-
LG14-3894	F7 LN97-15076 x LG01-7770	PI 483461	4767	33	116	125	2.8	1.2	0	12.7	-
LG14-3895	F7 LN97-15076 x LG01-7770	PI 483461	4517	33	113	120	2.6	1.0	0	11.5	-
LG14-3901	F7 LN97-15076 x LG01-7770	PI 483461	4610	35	119	126	3.1	1.2	0	12.1	-
LG14-3902	F7 LN97-15076 x LG01-7770	PI 483461	4256	30	114	131	3.6	1.9	0	13.1	-
LG14-3903	F7 LN97-15076 x LG01-7770	PI 483461	4418	31	115	124	3.1	1.6	0	12.9	-
LG12-9338	F7 Macon x PI 483461	PI 483461	4394	23	116	116	2.3	1.0	0	16.4	-
LG14-3029	F7 Williams 82 x PI 479752	PI 479752	3542	28	112	108	4.6	2.0	0	9.2	21
LG14-3030	F7 Williams 82 x PI 479752	PI 479752	3793	29	118	119	4.6	2.5	0	10.0	18
LG14-3044	F7 Williams 82 x PI 479752	PI 479752	3262	25	125	106	4.2	1.0	0	8.5	25
LG14-3051	F7 Williams 82 x PI 479752	PI 479752	3322	25	119	107	3.5	1.0	0	7.7	-
LG14-3053	F7 Williams 82 x PI 479752	PI 479752	3451	-	119	112	3.5	1.0	0	7.9	27
LG14-3070	F7 Williams 82 x PI 479752	PI 479753	3552	26	118	128	4.0	1.5	0	9.4	-
LG14-3088	F7 Williams 82 x PI 479752	PI 479752	3324	33	119	118	4.5	1.5	0	8.2	23
LG14-3173	F7 Williams 82 x PI 479752	PI 479752	1728	28	115	124	3.4	1.0	35	11.2	23
LG14-3175	F7 Williams 82 x PI 479752	PI 479752	3276	26	121	102	3.9	1.3	0	7.8	26
LG14-3178	F7 Williams 82 x PI 479752	PI 479752	3630	36	120	105	4.0	1.3	3	8.2	26
LG14-3179	F7 Williams 82 x PI 479752	PI 479752	3718	27	120	116	4.5	1.8	0	7.6	25
LG14-3183	F7 Williams 82 x PI 479752	PI 479752	3071	26	123	92	4.3	1.2	5	6.1	34
LG14-3185	F7 Williams 82 x PI 479752	PI 479752	3450	26	121	107	3.9	1.6	4	7.3	33
LG14-3188	F7 Williams 82 x PI 479752	PI 479752	3564	28	126	102	4.3	1.0	0	6.1	-
LG14-3191	F7 Williams 82 x PI 479752	PI 479752	3119	26	125	105	4.1	1.0	0	6.2	33
LG14-3198	F7 Williams 82 x PI 479752	PI 479752	3582	27	124	112	4.2	1.0	0	7.1	35
LG14-3211	F7 Williams 82 x PI 479752	PI 479752	3836	30	121	131	3.6	1.6	0	13.0	12
LG14-3218	F7 Williams 82 x PI 479752	PI 479752	3727	29	119	132	3.4	1.6	0	11.9	14
LG14-3220	F7 Williams 82 x PI 479752	PI 479752	4297	27	118	128	3.5	1.6	0	14.3	10
LSD ($p < 0.05$)			**269**	**1**	**1**	**7**	**0.3**	**0.2**	**4**	**0.7**	

Table 3.4. (cont.)

Entry	Pedigree	*G. soja* Parent	Yield (kg/ha)	R1 (days)	R8 (days)	Hgt† (cm)	Ldg‡ (1 - 5)	GH§ (1 - 5)	SH‖ (%)	SdWt†† (g)	% *G. soja*
LG14-3221	F7 Williams 82 x PI 479752	PI 479752	4064	26	117	130	3.7	1.6	0	12.8	10
LG14-3222	F7 Williams 82 x PI 479752	PI 479752	3984	27	118	119	3.7	1.5	0	12.5	9
LG14-3223	F7 Williams 82 x PI 479752	PI 479752	4035	27	118	130	3.6	1.7	0	13.9	9
LG14-3225	F7 Williams 82 x PI 479752	PI 479752	3830	28	117	117	3.3	1.4	0	12.9	16
LG14-3226	F7 Williams 82 x PI 479752	PI 479752	3924	27	114	116	3.3	1.4	1	13.0	-
LG14-3228	F7 Williams 82 x PI 479752	PI 479752	3741	24	114	119	4.0	2.0	0	12.3	15
LG14-3233	F7 Williams 82 x PI 479752	PI 479752	3891	28	114	126	3.1	1.4	0	13.5	12
LG14-3235	F7 Williams 82 x PI 479752	PI 479752	4178	26	116	121	3.4	1.5	0	14.3	13
LG14-3238	F7 Williams 82 x PI 479752	PI 479752	2981	23	113	114	3.2	1.4	0	6.7	33
LG14-3243	F7 Williams 82 x PI 479752	PI 479752	3041	22	112	105	3.5	1.4	0	7.5	38
LG14-3247	F7 Williams 82 x PI 479752	PI 479752	2754	22	113	119	3.7	1.1	0	7.5	37
LG14-3258	F7 Williams 82 x PI 479752	PI 479752	3316	23	116	107	3.9	1.4	0	7.8	35
LG14-3259	F7 Williams 82 x PI 479752	PI 479752	2852	23	113	108	3.6	1.4	0	6.5	36
LG14-3260	F7 Williams 82 x PI 479752	PI 479752	3224	-	119	106	3.4	1.0	0	7.1	36
LG14-3261	F7 Williams 82 x PI 479752	PI 479752	2852	27	117	112	3.5	1.4	0	6.5	37
LG14-3263	F7 Williams 82 x PI 479752	PI 479752	3053	23	112	99	3.6	1.7	0	6.4	41
LG14-3268	F7 Williams 82 x PI 479752	PI 479752	2938	24	111	119	4.4	1.8	0	6.1	42
LG14-3277	F7 Williams 82 x PI 479752	PI 479752	3392	23	113	94	2.9	1.1	0	7.0	31
LG14-3282	F7 Williams 82 x PI 479752	PI 479752	3700	23	113	104	3.1	1.1	0	9.1	-
LG14-3289	F7 Williams 82 x PI 479752	PI 479752	2918	23	114	100	3.7	1.3	0	6.9	37
LG14-3291	F7 Williams 82 x PI 479752	PI 479752	2671	23	109	103	3.6	1.6	0	6.6	39
LG14-3293	F7 Williams 82 x PI 479752	PI 479752	3221	23	114	111	3.3	1.2	0	6.6	37
LG14-3319	F7 Williams 82 x PI 479752	PI 479752	2867	33	113	117	4.1	2.1	0	7.3	25
LG14-3371	F7 Williams 82 x PI 479752	PI 479752	3478	26	118	102	3.4	1.2	0	9.0	25
LG14-3388	F7 Williams 82 x PI 479752	PI 479752	3575	24	123	122	3.7	1.0	11	7.4	23
LG14-3390	F7 Williams 82 x PI 479752	PI 479752	3536	24	116	122	4.0	1.8	0	9.9	33
LSD ($p < 0.05$)			**269**	**1**	**1**	**7**	**0.3**	**0.2**	**4**	**0.7**	

Table 3.4. (cont.)

Entry	Pedigree	*G. soja* Parent	Yield (kg/ha)	R1 (days)	R8 (days)	Hgt† (cm)	Ldg‡ (1 - 5)	GH§ (1 - 5)	SH⁋ (%)	SdWt†† (g)	% *G. soja*
LG14-3391	F7 Williams 82 x PI 479752	PI 479752	3397	25	123	135	4.5	2.3	0	9.6	28
LG14-3411	F7 Williams 82 x PI 479752	PI 479752	3758	34	117	120	3.7	1.7	0	11.7	20
LG14-3412	F7 Williams 82 x PI 479752	PI 479752	3642	37	116	112	3.6	1.8	0	8.7	22
LG14-3413	F7 Williams 82 x PI 479752	PI 479752	3654	35	115	111	3.4	1.9	0	8.5	24
LG14-3415	F7 Williams 82 x PI 479752	PI 479752	3641	37	117	123	3.4	1.7	0	10.0	20
LG14-3439	F7 Williams 82 x PI 479752	PI 479752	3698	36	117	119	3.5	1.7	0	8.4	20
LG14-3856	F8 SS98-3403 (3) x PI 483461	PI 483461	3315	27	117	116	2.6	1.0	0	10.3	-
LG14-3860	F8 SS98-3403 (3) x PI 483461	PI 483461	3605	24	118	120	2.7	1.0	0	9.9	-
LG14-3868	F8 SS98-3403 (3) x PI 483461	PI 483461	3818	26	119	124	3.0	1.0	0	10.0	-
LG14-3870	F8 SS98-3403 (3) x PI 483461	PI 483461	3534	26	118	123	2.6	1.0	0	9.9	-
LG14-3883	F8 SS98-3403 (3) x PI 483461	PI 483461	3356	30	121	118	3.4	1.0	0	10.4	-
LSD (p < 0.05)			**269**	**1**	**1**	**7**	**0.3**	**0.2**	**4**	**0.7**	

† Plant height

‡ Lodging

§ Growth Habit

⁋ Shattering

†† 100-Seed Weight

Table 3.5. Trait means for *G. soja*-derived lines and checks analyzed in MS IV test.

Entry	Pedigree	*G. soja* Parent	Yield (kg/ha)	R1 (days)	R8 (days)	Hgt† (cm)	Ldg‡ (1 - 5)	GH§ (1 - 5)	SH‖ (%)	SdWt†† (g)	% *G. soja*
LD06-7620	Check		5248	27	120	101	2.2	1.0	0	14.3	
LG04-6000	Check		5054	29	121	113	2.3	1.1	0	13.9	
LG11-6210	Check		5257	29	116	108	2.5	1.0	0	-	
LN97-15076	Check		4380	30	119	114	2.3	1.0	0	14.7	
LG14-3820	F5 LG04-5190 x PI 507788	PI 507788	3116	34	118	128	2.8	1.5	18	9.5	-
LG14-3821	F5 LG04-5190 x PI 507788	PI 507788	3533	28	118	107	2.7	1.0	5	12.0	-
LG14-3727	F5 LG04-6000 x PI 507787	PI 507787	3562	32	122	106	2.8	1.2	5	9.6	-
LG14-3728	F5 LG04-6000 x PI 507787	PI 507787	3760	30	123	110	3.2	1.4	1	9.6	22
LG14-3732	F5 LG04-6000 x PI 507787	PI 507787	2910	29	123	107	3.2	1.2	1	7.0	25
LG14-3733	F5 LG04-6000 x PI 507787	PI 507787	2909	41	125	109	3.3	1.1	30	6.6	24
LG14-3734	F5 LG04-6000 x PI 507787	PI 507787	2727	35	122	103	3.3	1.0	0	8.0	25
LG14-3735	F5 LG04-6000 x PI 507787	PI 507787	2502	32	124	109	3.7	1.4	0	7.2	26
LG14-3743	F5 LG04-6000 x PI 507787	PI 507787	3090	29	125	122	3.8	1.6	1	9.5	22
LG14-3744	F5 LG04-6000 x PI 507787	PI 507787	3063	29	120	113	3.3	1.6	0	11.1	-
LG14-3749	F5 LG04-6000 x PI 507787	PI 507787	3186	30	122	116	3.8	1.4	0	10.2	21
LG13-9751	F5 LG04-6000 x PI549048	PI 549048	3705	47	122	87	3.1	1.0	0	8.0	24
LG14-6069	F6 03JR309156 x LG07-2640	PI 483461	4463	25	118	118	2.7	1.0	0	13.1	-
LG14-6115	F6 03JR309156 x LG07-4727	PI 507807	4116	29	118	118	2.4	1.1	0	17.5	-
LG14-6141	F6 03JR309156 x LG07-4727	PI 507807	4230	29	119	108	2.1	1.0	0	15.7	-
LG14-6158	F6 03JR309156 x LG07-4727	PI 507807	4502	30	119	122	3.0	1.0	0	14.9	-
LG14-7929	F6 03JR309156 x LG07-4727	PI 507807	4273	26	119	122	3.3	1.2	0	17.2	-
LG14-7933	F6 03JR309156 x LG07-4727	PI 507807	4366	26	119	124	2.6	1.0	0	20.3	-
LG14-7936	F6 03JR309156 x LG07-4727	PI 507807	4501	29	119	124	2.9	1.0	0	17.6	-
LG14-7944	F6 03JR309156 x LG07-4727	PI 507807	4520	27	122	125	3.1	1.1	0	15.0	-
LSD ($p < 0.05$)			**255**	**2**	**1**	**6**	**0.3**	**0.2**	**5**	**0.6**	

Table 3.5. (cont.)

Entry	Pedigree	*G. soja* Parent	Yield (kg/ha)	R1 (days)	R8 (days)	Hgt† (cm)	Ldg‡ (1 - 5)	GH§ (1 - 5)	SH⁋ (%)	SdWt†† (g)	% *G. soja*
LG14-7946	F6 03JR309156 x LG07-4727	PI 507807	4629	30	124	123	3.2	1.0	0	16.2	-
LG14-7947	F6 03JR309156 x LG07-4727	PI 507807	4340	29	122	126	3.5	1.4	0	15.6	-
LG14-7948	F6 03JR309156 x LG07-4727	PI 507807	4765	27	121	123	3.2	1.3	0	15.9	-
LG14-7949	F6 03JR309156 x LG07-4727	PI 507807	4574	29	123	124	2.9	1.1	0	16.2	-
LG14-7952	F6 03JR309156 x LG07-4727	PI 507807	4476	29	122	130	3.3	1.5	0	15.6	-
LG14-2570	F6 IA3023 x PI522183A	PI 522183A	2442	49	118	78	3.6	1.1	1	9.7	26
LG14-2571	F6 IA3023 x PI522183A	PI 522183A	2459	61	128	96	3.5	1.0	0	9.0	23
LG14-2651	F6 IA3023 x PI522183A	PI 522183A	2839	52	118	109	4.2	1.9	0	7.0	18
LG14-2780	F6 IA3023 x PI549048	PI 549048	2664	51	124	123	4.0	1.9	16	9.7	23
LG14-2782	F6 IA3023 x PI549048	PI 549048	2589	53	131	130	4.0	1.6	26	9.4	25
LG14-2785	F6 IA3023 x PI549048	PI 549048	2939	53	130	131	3.5	1.6	20	9.5	23
LG14-2790	F6 IA3023 x PI549048	PI 549048	3024	52	130	126	3.4	1.6	5	9.7	22
LG14-2799	F6 IA3023 x PI549048	PI 549048	3480	32	127	124	3.4	1.5	11	10.3	20
LG14-2800	F6 IA3023 x PI549048	PI 549048	3128	33	128	121	2.9	1.6	23	10.2	22
LG14-2802	F6 IA3023 x PI549048	PI 549048	3730	30	123	110	3.9	1.7	1	9.3	20
LG14-3486	F6 LG04-6000 x PI549048	PI 549048	3228	45	120	102	2.8	1.1	5	7.0	24
LG14-3509	F6 LG04-6000 x PI549048	PI 549048	3059	50	118	107	3.6	1.2	0	7.3	25
LG14-3511	F6 LG04-6000 x PI549048	PI 549048	3144	50	119	108	4.1	1.3	0	6.5	24
LG14-3514	F6 LG04-6000 x PI549048	PI 549048	2868	52	123	122	3.8	1.3	36	7.8	28
LG14-3567	F6 LG04-6000 x PI549048	PI 549048	2829	31	119	109	2.9	1.1	29	7.8	25
LG14-3582	F6 LG04-6000 x PI549048	PI 549048	3065	52	119	113	3.5	1.4	0	8.4	26
LG14-3587	F6 LG04-6000 x PI549048	PI 549048	2967	53	125	116	3.3	1.1	0	6.7	29
LG14-3589	F6 LG04-6000 x PI549048	PI 549048	3131	52	127	118	3.3	1.1	29	7.5	29
LG14-3592	F6 LG04-6000 x PI549048	PI 549048	2871	48	124	116	3.2	1.3	18	8.1	26
LSD (p < 0.05)			**255**	**2**	**1**	**6**	**0.3**	**0.2**	**5**	**0.6**	

Table 3.5. (cont.)

Entry	Pedigree	*G. soja* Parent	Yield (kg/ha)	R1 (days)	R8 (days)	Hgt† (cm)	Ldg‡ (1 - 5)	GH§ (1 - 5)	SH⁋ (%)	SdWt†† (g)	% *G. soja*
LG14-3600	F6 LG04-6000 x PI549048	PI 549048	2907	49	120	116	4.0	1.0	0	8.3	23
LG14-3603	F6 LG04-6000 x PI549048	PI 549048	3342	28	118	119	3.4	1.1	0	7.1	-
LG14-3608	F6 LG04-6000 x PI549048	PI 549048	2625	51	120	115	3.9	1.8	38	7.4	29
LG14-3622	F6 LG04-6000 x PI549048	PI 549048	3172	47	119	106	4.0	1.3	0	6.6	28
LG14-3639	F6 LG04-6000 x PI549048	PI 549048	2815	25	123	110	3.9	1.4	0	7.7	32
LG14-3648	F6 LG04-6000 x PI549048	PI 549048	2954	27	124	116	3.9	1.6	0	9.7	34
LG14-3654	F6 LG04-6000 x PI549048	PI 549048	2758	27	122	114	3.9	1.1	0	7.2	28
LG14-3675	F6 LG04-6000 x PI549048	PI 549048	2728	28	121	116	3.5	1.1	0	7.2	32
LG14-3679	F6 LG04-6000 x PI549048	PI 549048	2835	27	124	118	3.4	1.1	0	7.3	32
LG14-3683	F6 LG04-6000 x PI549048	PI 549048	2448	26	123	123	3.8	1.3	0	6.8	29
LG14-3684	F6 LG04-6000 x PI549048	PI 549048	3063	29	126	118	3.8	1.5	0	7.8	-
LG14-3686	F6 LG04-6000 x PI549048	PI 549048	2500	26	125	114	4.0	1.4	0	8.4	31
LG14-3688	F6 LG04-6000 x PI549048	PI 549048	2674	25	127	107	3.6	1.3	0	7.7	32
LG14-3689	F6 LG04-6000 x PI549048	PI 549048	2895	29	128	125	4.1	1.3	0	6.8	31
LG14-3691	F6 LG04-6000 x PI549048	PI 549048	2631	26	129	120	3.3	1.3	0	7.1	28
LG14-3719	F6 LG04-6000 x PI549048	PI 549048	2995	47	120	89	2.7	1.2	0	7.4	25
LG14-2854	F6 LG97-7012 x PI468399A	PI 468399A	2647	44	122	104	4.3	1.6	0	6.1	-
LG08-4152	F6 Williams 82 x LG01-7728	PI 479767	3565	36	121	116	3.0	1.1	0	10.8	-
LG14-3897	F7 LN97-15076 x LG01-7770	PI 483461	4539	35	124	117	2.5	1.0	0	12.3	-
LG14-3899	F7 LN97-15076 x LG01-7770	PI 483461	4533	34	120	116	2.6	1.1	0	13.4	-
LG14-2971	F7 Williams 82 x PI 479752	PI 479752	2709	33	121	122	4.3	1.5	0	9.9	17
LG14-2973	F7 Williams 82 x PI 479752	PI 479752	3314	29	123	103	4.6	1.8	0	10.2	18
LG14-2974	F7 Williams 82 x PI 479752	PI 479752	3034	30	123	138	4.5	1.5	0	9.8	-
LG14-2976	F7 Williams 82 x PI 479752	PI 479752	3319	29	122	122	4.1	1.3	0	10.1	17
LSD ($p < 0.05$)			**255**	**2**	**1**	**6**	**0.3**	**0.2**	**5**	**0.6**	

Table 3.5. (cont.)

Entry	Pedigree	*G. soja* Parent	Yield (kg/ha)	R1 (days)	R8 (days)	Hgt† (cm)	Ldg‡ (1 - 5)	GH§ (1 - 5)	SH⁋ (%)	SdWt†† (g)	% *G. soja*
LG14-3000	F7 Williams 82 x PI 479752	PI 479752	2737	29	130	150	4.3	1.8	0	9.8	19
LG14-3032	F7 Williams 82 x PI 479752	PI 479752	3228	28	123	117	3.3	1.2	0	10.4	22
LG14-3039	F7 Williams 82 x PI 479752	PI 479752	3392	34	125	110	4.2	1.0	0	8.9	25
LG14-3041	F7 Williams 82 x PI 479752	PI 479752	3725	31	125	117	4.0	1.4	0	8.6	26
LG14-3046	F7 Williams 82 x PI 479752	PI 479752	3500	37	132	115	3.4	1.1	0	11.1	24
LG14-3049	F7 Williams 82 x PI 479752	PI 479752	3552	31	127	112	3.3	1.3	0	9.5	24
LG14-3050	F7 Williams 82 x PI 479752	PI 479752	3428	31	124	111	3.4	1.0	0	8.1	26
LG14-3052	F7 Williams 82 x PI 479752	PI 479752	3484	31	124	114	3.5	1.0	0	8.4	26
LG14-3055	F7 Williams 82 x PI 479752	PI 479752	3499	30	124	113	3.6	1.0	0	7.9	-
LG14-3056	F7 Williams 82 x PI 479752	PI 479752	3326	29	125	116	3.5	1.0	0	8.9	26
LG14-3057	F7 Williams 82 x PI 479752	PI 479752	3282	31	124	113	3.3	1.1	0	8.0	26
LG14-3058	F7 Williams 82 x PI 479752	PI 479752	3494	28	123	116	3.3	1.0	0	8.0	27
LG14-3076	F7 Williams 82 x PI 479752	PI 479752	3952	32	125	129	3.4	1.8	0	10.5	22
LG14-3094	F7 Williams 82 x PI 479752	PI 479752	3306	28	121	107	4.0	1.0	7	7.5	31
LG14-3100	F7 Williams 82 x PI 479752	PI 479752	3413	52	130	111	4.0	1.3	0	8.0	26
LG14-3101	F7 Williams 82 x PI 479752	PI 479752	3631	53	128	108	3.8	1.3	1	7.6	26
LG14-3103	F7 Williams 82 x PI 479752	PI 479752	3412	52	129	109	3.8	1.1	0	8.1	26
LG14-3106	F7 Williams 82 x PI 479752	PI 479752	3512	50	127	100	3.7	1.0	0	8.6	26
LG14-3118	F7 Williams 82 x PI 479752	PI 479752	2761	52	121	125	4.5	1.4	0	8.6	23
LG14-3122	F7 Williams 82 x PI 479752	PI 479752	2888	50	120	117	4.7	1.5	0	8.7	22
LG14-3159	F7 Williams 82 x PI 479752	PI 479752	2795	27	120	116	4.1	1.6	0	6.9	28
LG14-3160	F7 Williams 82 x PI 479752	PI 479752	2915	29	120	114	4.0	1.6	0	6.5	28
LG14-3161	F7 Williams 82 x PI 479752	PI 479752	2704	47	118	115	3.8	1.8	1	6.9	25
LG14-3164	F7 Williams 82 x PI 479752	PI 479752	3038	28	119	114	3.9	1.6	0	7.3	29
LSD ($p < 0.05$)			**255**	**2**	**1**	**6**	**0.3**	**0.2**	**5**	**0.6**	

Table 3.5. (cont.)

Entry	Pedigree	*G. soja* Parent	Yield (kg/ha)	R1 (days)	R8 (days)	Hgt† (cm)	Ldg‡ (1 - 5)	GH§ (1 - 5)	SH‖ (%)	SdWt†† (g)	% *G. soja*
LG14-3166	F7 Williams 82 x PI 479752	PI 479752	3312	28	120	112	4.3	1.4	0	7.7	31
LG14-3169	F7 Williams 82 x PI 479752	PI 479752	943	30	114	117	3.6	1.6	48	10.0	27
LG14-3176	F7 Williams 82 x PI 479752	PI 479752	3542	36	125	109	4.2	1.9	2	7.7	26
LG14-3192	F7 Williams 82 x PI 479752	PI 479752	3249	27	128	106	4.0	1.1	3	6.3	34
LG14-3194	F7 Williams 82 x PI 479752	PI 479752	3216	34	126	104	4.4	1.4	0	6.3	34
LG14-3200	F7 Williams 82 x PI 479752	PI 479752	3833	28	119	106	3.3	1.5	0	7.1	31
LG14-3201	F7 Williams 82 x PI 479752	PI 479752	3592	28	117	104	3.0	1.5	0	6.7	32
LG14-3213	F7 Williams 82 x PI 479752	PI 479752	3598	34	120	125	3.5	1.8	0	12.8	13
LG14-3215	F7 Williams 82 x PI 479752	PI 479752	3569	31	117	119	3.9	2.0	0	13.3	13
LG14-3217	F7 Williams 82 x PI 479752	PI 479752	3566	31	121	137	2.6	1.6	0	12.5	13
LG14-3245	F7 Williams 82 x PI 479752	PI 479752	2773	27	124	126	3.3	1.5	0	7.6	39
LG14-3266	F7 Williams 82 x PI 479752	PI 479752	3064	29	131	131	3.9	1.8	0	6.8	37
LG14-3287	F7 Williams 82 x PI 479752	PI 479752	2704	25	125	114	3.1	1.3	0	7.2	-
LG14-3323	F7 Williams 82 x PI 479752	PI 479752	3023	49	121	123	3.6	1.5	0	8.8	36
LG14-3365	F7 Williams 82 x PI 479752	PI 479752	3930	28	121	113	3.3	1.3	0	9.0	33
LG14-3366	F7 Williams 82 x PI 479752	PI 479752	3231	26	118	106	3.3	1.3	0	9.0	33
LG14-3367	F7 Williams 82 x PI 479752	PI 479752	3138	26	120	110	3.7	1.2	0	9.2	32
LG14-3373	F7 Williams 82 x PI 479752	PI 479752	2717	27	132	134	3.9	1.5	8	7.1	31
LG14-3374	F7 Williams 82 x PI 479752	PI 479752	2494	28	122	125	3.9	1.6	23	7.4	-
LG14-3378	F7 Williams 82 x PI 479752	PI 479752	3432	24	123	124	3.3	1.3	4	8.4	-
LG14-3398	F7 Williams 82 x PI 479752	PI 479752	3562	28	116	118	3.4	1.4	2	9.5	30
LG14-3417	F7 Williams 82 x PI 479752	PI 479752	3540	38	121	113	3.3	1.5	0	10.4	23
LG14-3419	F7 Williams 82 x PI 479752	PI 479752	3503	35	118	110	3.1	1.5	0	9.7	23
LG14-3421	F7 Williams 82 x PI 479752	PI 479752	3716	36	121	115	3.3	1.5	0	11.0	-
LSD ($p < 0.05$)			**255**	**2**	**1**	**6**	**0.3**	**0.2**	**5**	**0.6**	

Table 3.5. (cont.)

Entry	Pedigree	*G. soja* Parent	Yield (kg/ha)	R1 (days)	R8 (days)	Hgt† (cm)	Ldg‡ (1 - 5)	GH§ (1 - 5)	SH⁋ (%)	SdWt†† (g)	% *G. soja*
LG14-3422	F7 Williams 82 x PI 479752	PI 479752	3546	37	121	115	3.2	1.3	0	10.1	23
LG14-3425	F7 Williams 82 x PI 479752	PI 479752	3594	33	120	114	4.0	1.6	0	9.6	25
LG14-3426	F7 Williams 82 x PI 479752	PI 479752	3593	33	117	110	3.3	1.7	0	9.0	25
LG14-3444	F7 Williams 82 x PI 479752	PI 479752	3553	35	121	122	3.1	1.7	0	8.8	20
LG14-3446	F7 Williams 82 x PI 479752	PI 479752	3478	36	119	123	3.4	1.7	0	9.2	21
LG14-3861	F8 SS98-3403 (3) x PI 483461	PI 483461	3592	44	119	124	2.9	1.0	1	9.9	-
LG14-3866	F8 SS98-3403 (3) x PI 483461	PI 483461	3366	45	118	130	3.0	1.3	0	10.2	-
LSD (p < 0.05)			**255**	**2**	**1**	**6**	**0.3**	**0.2**	**5**	**0.6**	

† Plant height

‡ Lodging

§ Growth Habit

⁋ Shattering

†† 100-Seed Weight

Table 3.6. Yield performance for each cross. Only crosses represented by at least two lines are included.

Cross	BC†	Avg. Yield (kg/ha)	Yield Range (kg/ha)	Total Lines
03JR309156 x LG07-2640	3	4783	4332 - 5295	31
06NB204846 x LG07-4307	2	4542	4107 - 5192	26
03JR309156 x LG07-4727	2	4522	4116 - 4918	43
LN97-15076 x LG01-7770	2	4520	4256 - 4767	7
06NB204846 x LG07-2526	3	4366	3828 - 4789	25
IA2052 x LG01-7812	2	4297	4077 - 4524	4
Macon x PI 483461	0	4157	3920 - 4394	2
Williams 82 (2) x PI 479767	1	3988	3893 - 4138	3
06NB204846 x LG07-4216	3	3875	3516 - 4114	11
SS98-3403 (3) x PI 483461	2	3512	3315 - 3818	7
Williams 82 x PI 479752	0	3317	739 - 4297	123
LG04-6000 x PI 507787	0	3262	2502 - 3914	22
LG04-5190 x PI 507788	0	3122	2472 - 3587	17
LG04-6000 x PI549048	0	3059	2448 - 3705	47
IA3023 x PI424068	0	3041	3036 - 3046	2
IA3023 x PI549048	0	3039	2589 - 3730	11
IA3023 x PI468397A	0	2902	2228 - 3516	10
LG97-7012 x PI468399A	0	2475	1548 - 3082	4
IA3023 x PI522183A	0	2446	1023 - 3290	15

† Number of backcrosses to *G. max*

Table 3.7. *G. soja*-derived lines from MS IV with LN97-15076 as a parent with yields equivalent to LN97-15076 at $\alpha = 0.05$.

Entry	Pedigree	Yield (kg/ha)	R1†	R8‡	Hgt§ (cm)	Ldg⁋ (1 - 5)	GH†† (1 - 5)	SH‡‡ (%)	SdWt§§ (g)
LG14-3897	F7 LN97-15076 x LG01-7770	4539	35	124	117	2.5	1.0	0	12.3
LG14-3899	F7 LN97-15076 x LG01-7770	4533	34	120	116	2.6	1.1	0	13.4
LN97-15076		4380	30	119	114	2.3	1.0	0	14.7
LSD (p < 0.05)		**255**	**2**	**1**	**6**	**0.3**	**0.2**	**5**	**0.6**

† Days after May 31 to first flowering

‡ Days after May 31 to maturity

§ Plant height

⁋ Lodging

†† Growth Habit

‡‡ Shattering

§§ 100-Seed Weight

Table 3.8. Selected *G. soja*-derived lines from MS III with Williams 82 as a parent that are equal or greater in yield to Williams 82 at α = 0.05.

Entry	Pedigree	Yield (kg/ha)	R1†	R8‡	Hgt§ (cm)	Ldg⁋ (1 - 5)	GH†† (1 - 5)	SH‡‡ (%)	SdWt§§ (g)
LG14-3220	F7 Williams 82 x PI 479752	4297*	27	118	128	3.5	1.6	0	14.3
LG12-9591	F5 Williams 82 x PI 479752	4292*	25	117	111	2.5	1.0	0	12.5
LG14-3235	F7 Williams 82 x PI 479752	4178	26	116	121	3.4	1.5	0	14.3
LG07-4231	F5 Williams 82 x LG01-7909	4143	29	117	134	3.2	1.3	0	14.1
LG11-9189	F5 Williams 82 (2) x PI 479767	4138	28	115	120	3.0	1.3	0	13.5
LG12-9619	F5 Williams 82 x PI 479752	4081	24	117	117	2.5	1.0	0	12.3
LG14-3221	F7 Williams 82 x PI 479752	4064	26	117	130	3.7	1.6	0	12.8
LG14-3223	F7 Williams 82 x PI 479752	4035	27	118	130	3.6	1.7	0	13.9
LG14-3222	F7 Williams 82 x PI 479752	3984	27	118	119	3.7	1.5	0	12.5
Williams 82		3954	28	118	128	2.9	1.2	0	16.1
LG11-9161	F5 Williams 82 (2) x PI 479767	3934	25	113	123	3.0	1.3	0	14.2
LG14-3226	F7 Williams 82 x PI 479752	3924	27	114	116	3.3	1.4	1	13.0
LG11-9166	F5 Williams 82 (2) x PI 479767	3893	30	115	118	3.0	1.1	0	12.5
LG14-3233	F7 Williams 82 x PI 479752	3891	28	114	126	3.1	1.4	0	13.5
LG14-3211	F7 Williams 82 x PI 479752	3836	30	121	131	3.6	1.6	0	13.0
LG14-3225	F7 Williams 82 x PI 479752	3830	28	117	117	3.3	1.4	0	12.9
LG14-3030	F7 Williams 82 x PI 479752	3793	29	118	119	4.6	2.5	0	10.0
LG14-3411	F7 Williams 82 x PI 479752	3758	34	117	120	3.7	1.7	0	11.7
LG14-3228	F7 Williams 82 x PI 479752	3741	24	114	119	4.0	2.0	0	12.3
LG14-3218	F7 Williams 82 x PI 479752	3727	29	119	132	3.4	1.6	0	11.9
LG14-3179	F7 Williams 82 x PI 479752	3718	27	120	116	4.5	1.8	0	7.6
LG14-3282	F7 Williams 82 x PI 479752	3700	23	113	104	3.1	1.1	0	9.1
LG14-3439	F7 Williams 82 x PI 479752	3698	36	117	119	3.5	1.7	0	8.4
LSD (p < 0.05)		**269**	**1**	**1**	**7**	**0.3**	**0.2**	**4**	**0.7**

* Significant difference from Williams 82 at α=0.05

† Days after May 31 to first flowering

‡ Days after May 31 to maturity

§ Plant height

⁋ Lodging

†† Growth Habit

‡‡ Shattering

§§ 100-Seed Weight

Table 3.9. Total contribution from *G. soja* measured by SNPs. The average and range of percent *G. soja* alleles is represented.

			% *G. soja* SNPs	
Cross	**BC**[†]	**Total Lines**	**Average**[††]	**Range**
IA3023 x PI 468397A	0	10	16.6***	12.4 - 22.8
IA3023 x PI 522183A	0	14	21.9***	17 - 26.8
IA3023 x PI 549048	0	10	21.1***	17.4 - 24.5
LG04-6000 x PI 507787	0	18	23.9***	17 - 27.2
LG04-6000 x PI 549048	0	43	26.9***	20.5 - 33.8
LN97-15076 x LG01-7770	2	7	7.6	7 - 8.7
SS98-3403 (3) x PI 483461	2	7	12.3	9.7 - 16.8
Williams 82 (2) x PI 479767	1	3	3.8**	3.3 - 4.2
Williams 82 x PI 479752	0	111	25.8***	8.8 - 41.7

** Significance at 0.01

*** Significance at < 0.0001

† Number of backcrosses to *G. max*

††Significant differences noted from what would be expected without selection.

Table 3.10. Large seeded Williams 82 x PI 479752 lines containing *G. soja* alleles at 100-seed weight QTL. The three QTL individually explained the largest phenotypic variance in the Williams 82 x PI 479752 mapping population.

	100-Seed Weight QTL[†]			SdWt[‡]	Yield	
Entry	***qSW-12***	***qSW-17-1***	***qSW-19***	**(g)**	**(kg/ha)**	**Test**
LG14-3030	S	M	S	10.0	3793	MS III
LG14-3415	S	S	M	10.0	3641	MS III
LG14-3215	S	M	M	13.3	3569	MS IV
LG14-3228	S	M	M	12.3	3741	MS III
LG14-3411	M	S	M	11.7	3758	MS III
LG14-3417	M	S	M	10.4	3540	MS IV
LG14-3327	S	M	M	10.2	2326	MS II
LG14-3422	M	S	M	10.1	3546	MS IV
Williams 82	M	M	M	16.1	3954	MS III

[†] S indicates loci homozygous for *G. soja* allele, M indicates loci homozygous for *G. max* allele

[‡] 100-seed weight

FIGURES

Figure 3.1. Distribution of *G. soja* allele frequency across all 20 chromosomes in 111 lines from Williams 82 x PI 479752. Grey bars at bottom indicate the location of 32 regions containing domestication-related QTL. Green horizontal line indicates the expected allele frequency of 0.5 in a biparental population.

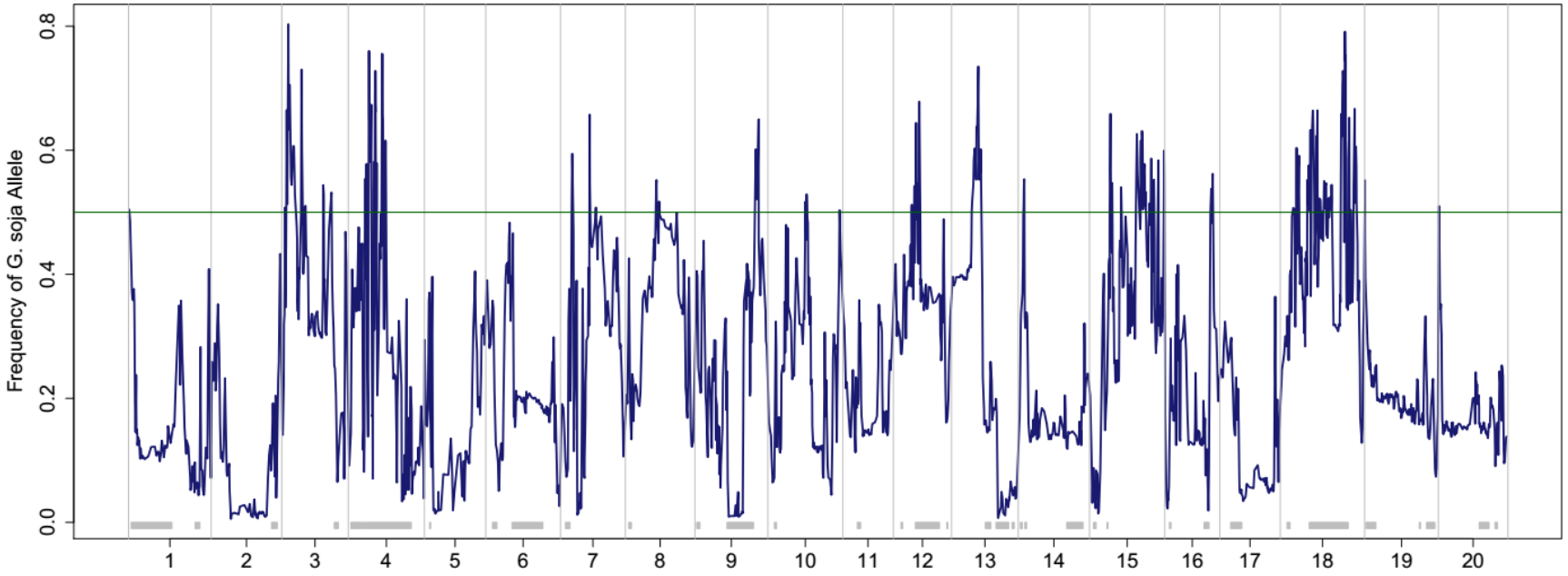


Figure 3.2. Graph of average yield in MS III versus percentage of *G. soja* alleles for lines from LG04-6000 x PI 549048 and Williams 82 x PI 479752. The number of lines within each pedigree is given in parentheses, and the slope of the regression line is given following the semi-colon. Regression coefficients that are not significant are listed as 'NS'.

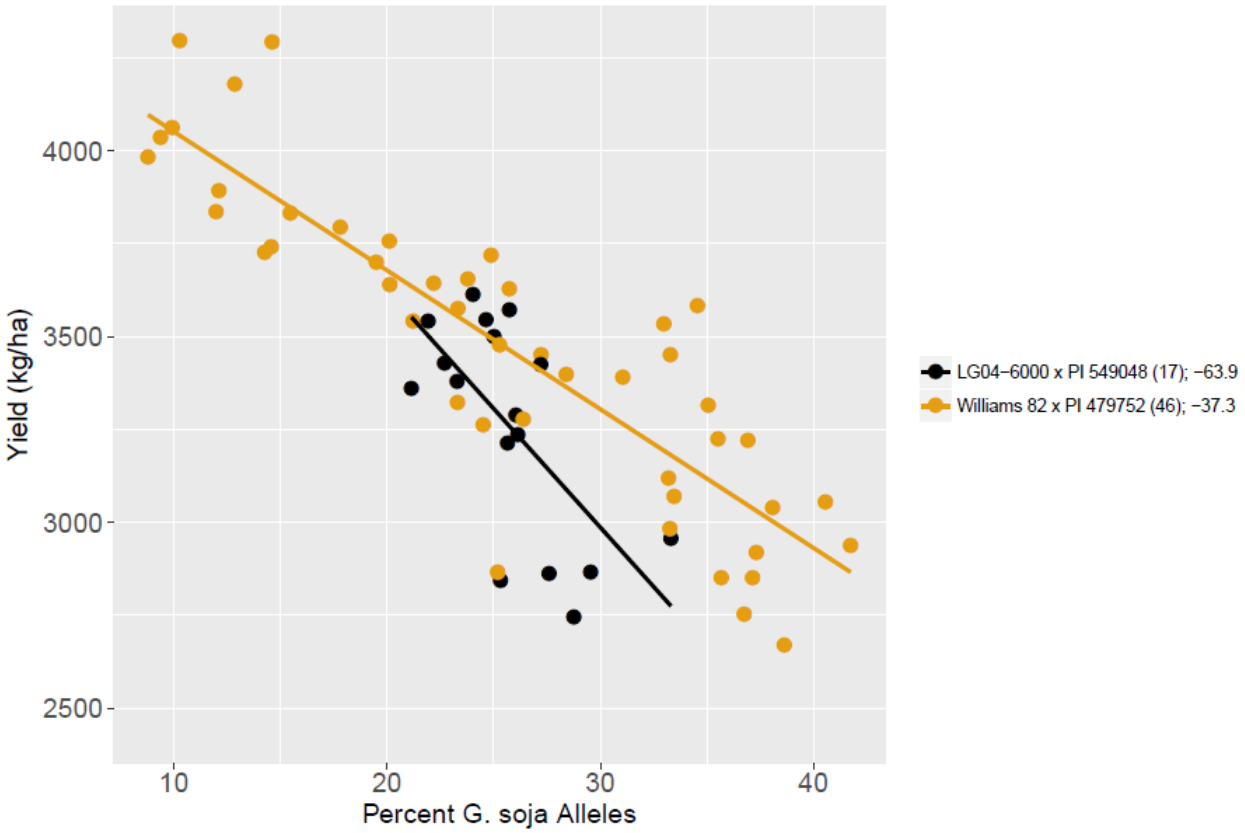


Figure 3.3. Graph of average yield in MS IV versus percentage of *G. soja* alleles for lines from LG04-6000 x PI 549048 and Williams 82 x PI 479752. The number of lines within each pedigree is given in parentheses, and the slope of the regression line is given following the semi-colon. Regression coefficients that are not significant are listed as 'NS'.

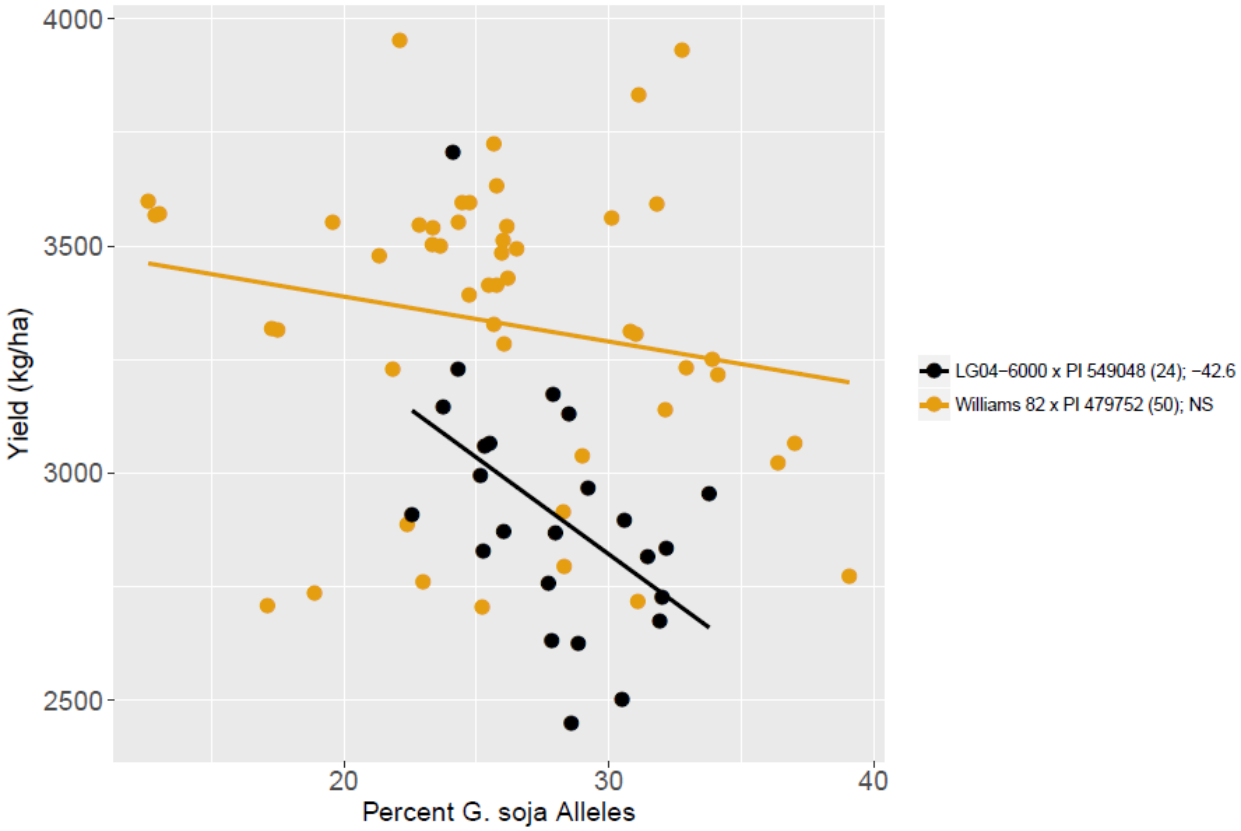


Figure 3.4. Graph of average 100-seed weight in MS III versus percentage of *G. soja* alleles for lines from LG04-6000 x PI 549048 and Williams 82 x PI 479752. The number of lines within each pedigree is given in parentheses, and the slope of the regression line is given following the semi-colon. Regression coefficients that are not significant are listed as 'NS'.

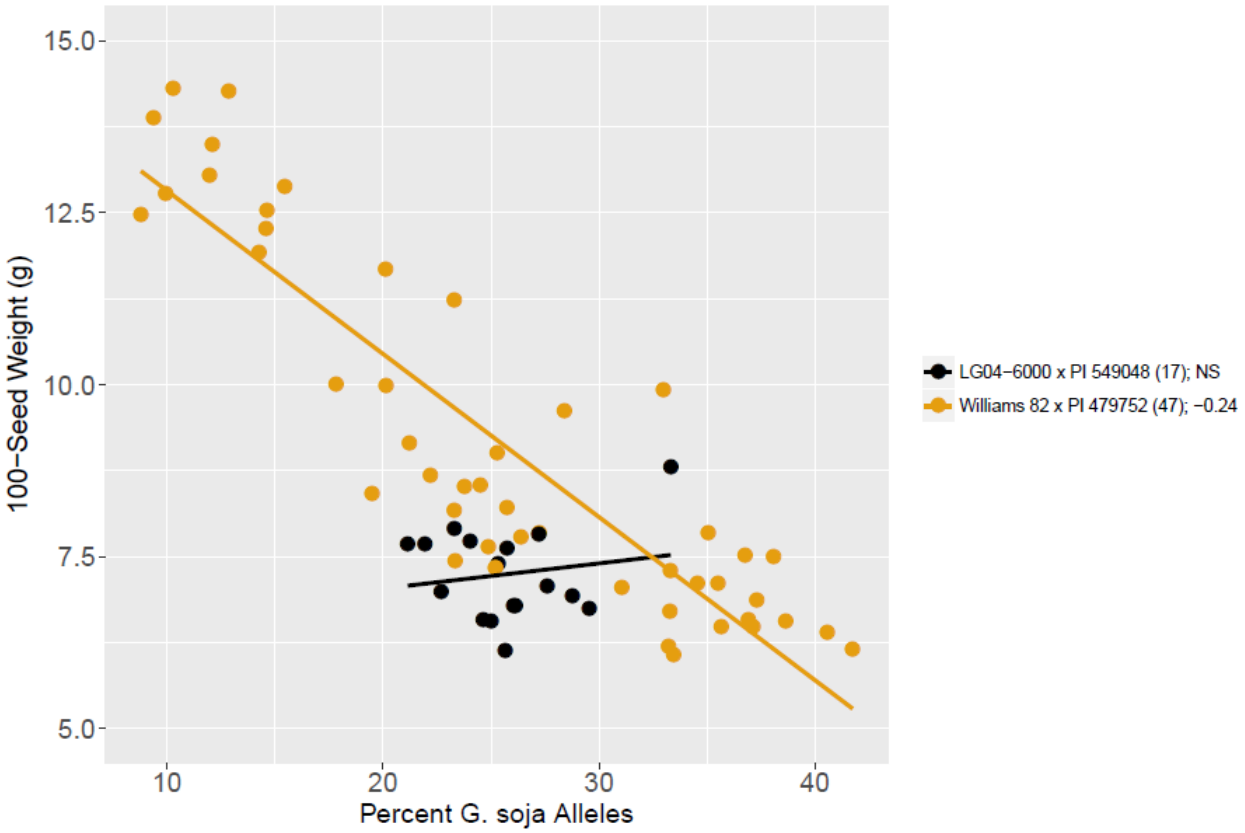


Figure 3.5. Graph of average 100-seed weight in MS IV versus percentage of *G. soja* alleles for lines from LG04-6000 x PI 549048 and Williams 82 x PI 479752. The number of lines within each pedigree is given in parentheses, and the slope of the regression line is given following the semi-colon. Regression coefficients that are not significant are listed as 'NS'.

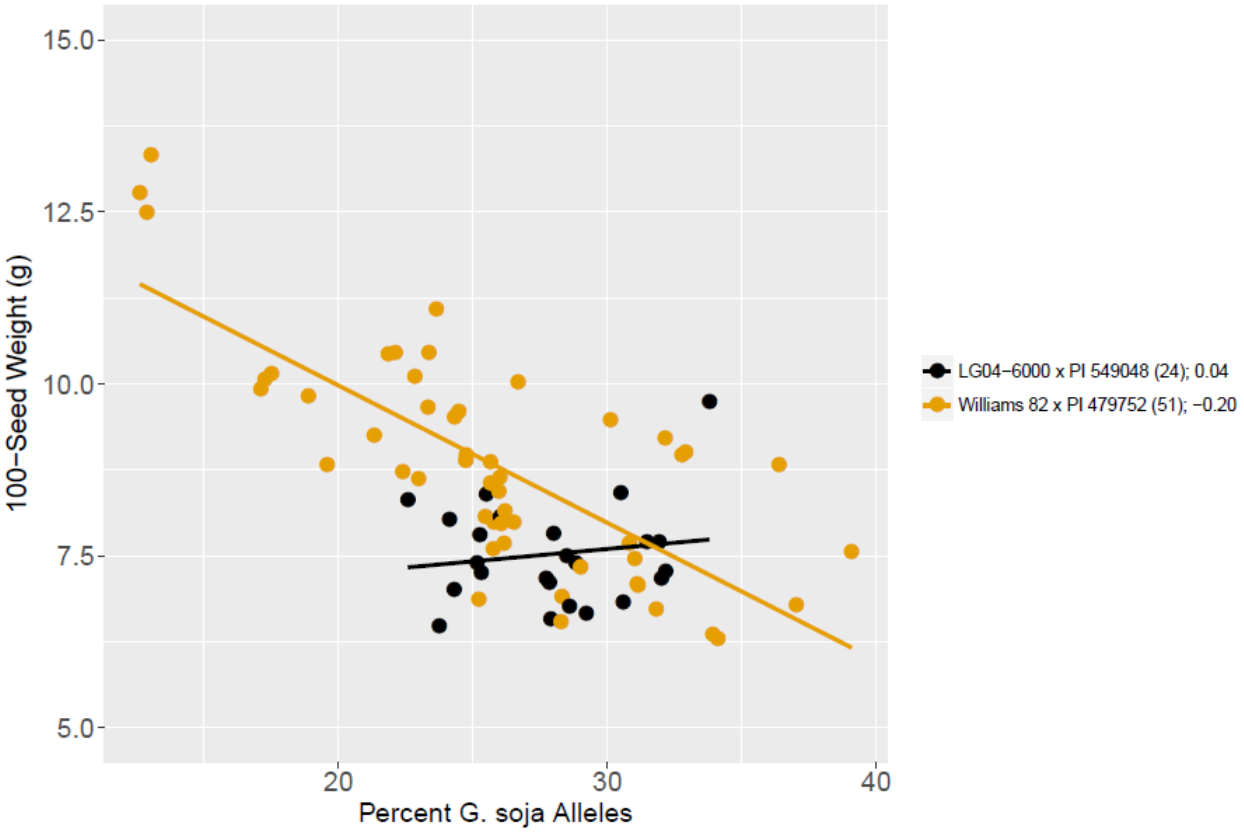


Figure 3.6. Frequency of *G. soja* alleles surrounding major shattering QTL (*qSH-16*) on chromosome 16. The location of the QTL is indicated by the grey bar.

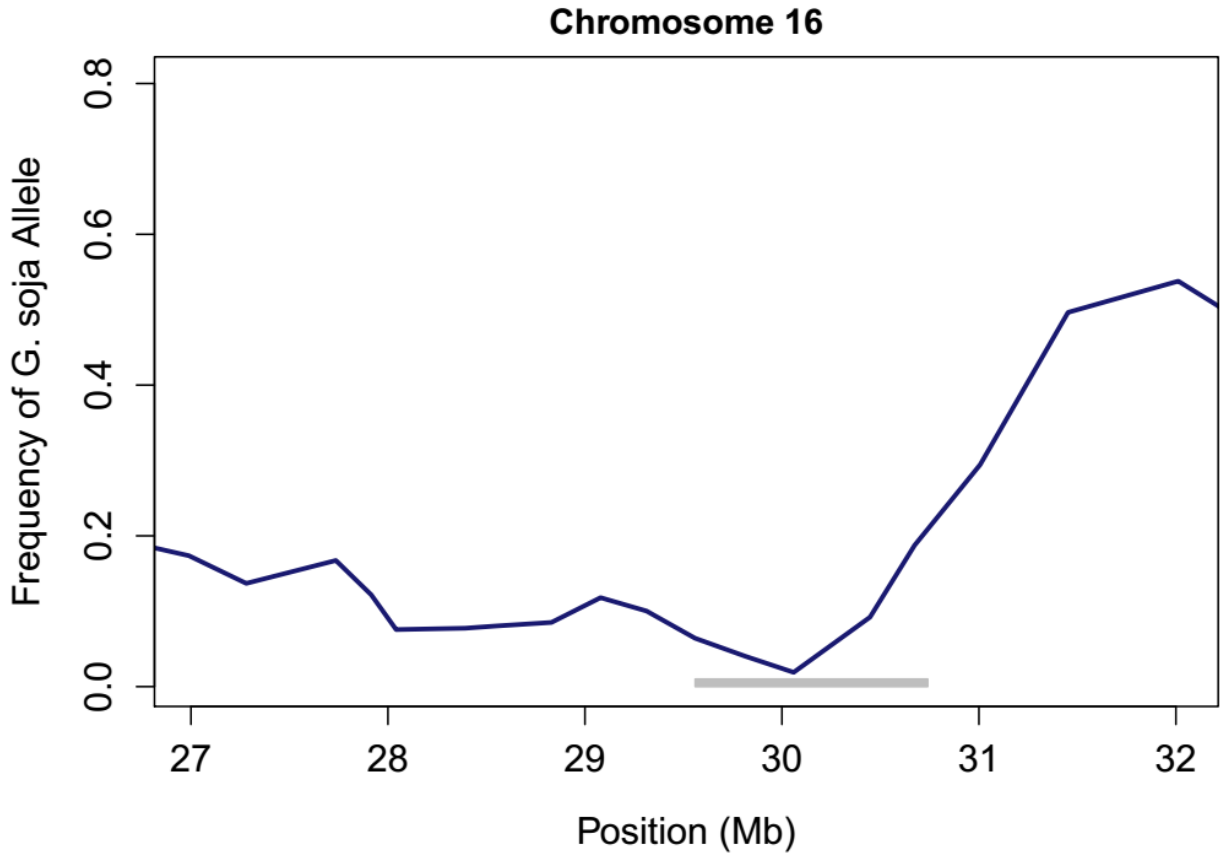
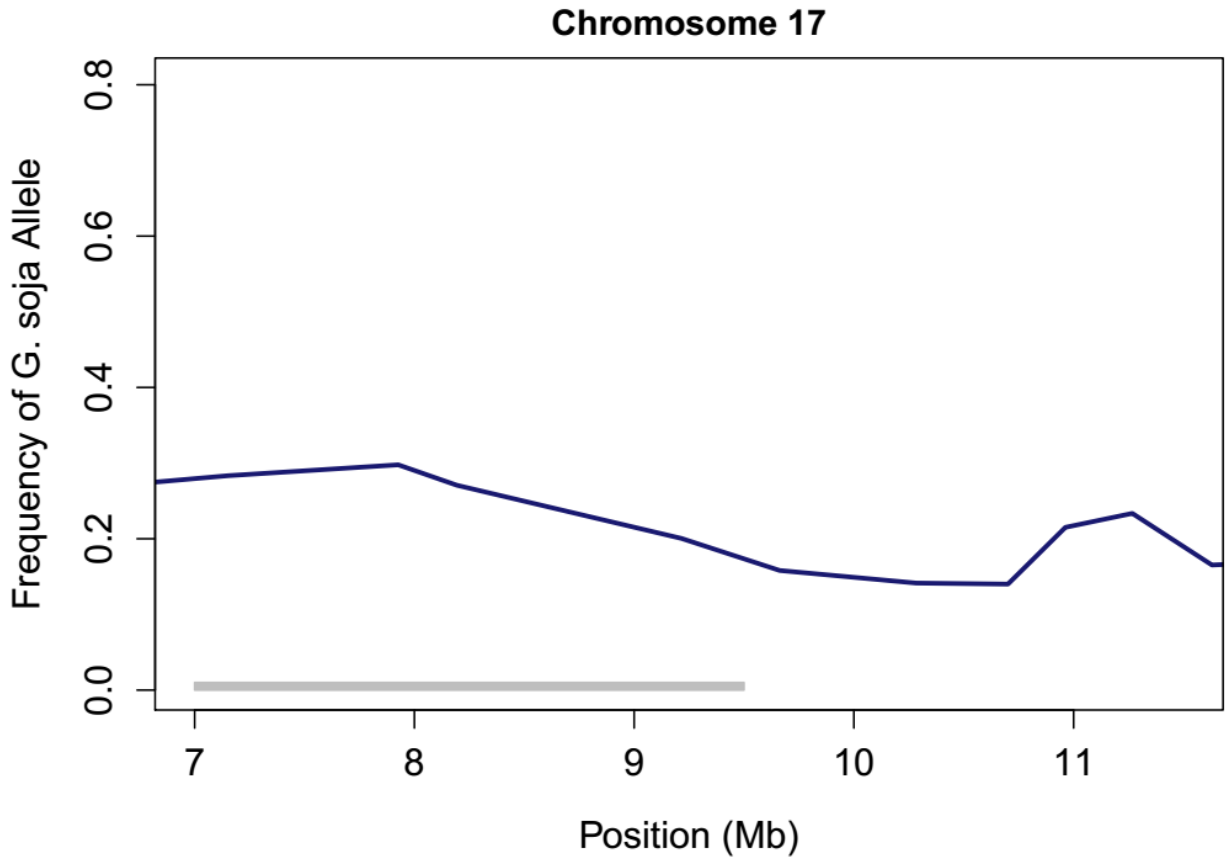


Figure 3.7. Frequency of *G. soja* alleles surrounding 100-seed weight QTL (*qSW-17-1*) on chromosome 17. The location of the QTL is indicated by the grey bar.



REFERENCES

Abdel-Haleem, H., H.R. Boerma, T.E. Carter, E.D. Wood, and Z. Li. 2015. Registration of G07-6012 and G07-6029 Soybean Germplasm, Which Derive 50% Pedigree from Wild Soybean. J. Plant Regist. 9(2): 222.

Akpertey, A., M. Belaffif, G.L. Graef, M.A. Rouf Mian, J. Grover Shannon, P.B. Cregan, M.E. Hudson, B.W. Diers, and R.L. Nelson. 2014. Effects of selective genetic introgression from wild soybean to soybean. Crop Sci. 54(6): 2683–2695.

Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67(1): 1–48.

Browning, B.L., and S.R. Browning. 2016. Genotype imputation with millions of reference samples. Am. J. Hum. Genet. 98(1): 116–126.

Carpenter, J.A., and W.R. Fehr. 1986. Genetic variability for desirable agronomic traits in populations containing *Glycine soja* germplasm. Crop Sci. 26(4): 681.

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(4): 575–582.

Cui, Z., A.T. James, S. Miyazaki, R.F. Wilson, and T.E. Carter. 2004. Breeding specialty soybeans for traditional and new soyfoods. p. 264–322. *In* Liu, K. (ed.), Soybeans as functional foods and ingredients. AOCS Press, Champaign, IL.

Diers, B.W., P.R. Arelli, S.R. Carlson, W.R. Fehr, E.A. Kabelka, R.C. Shoemaker, and D. Wang. 2005. Registration of LDX01-1-65 soybean germplasm with soybean cyst nematode resistance derived from *Glycine soja*. Crop Sci. 45(4): 1671–1672.

Dong, Y., X. Yang, J. Liu, B.H. Wang, B.L. Liu, and Y.Z. Wang. 2014. Pod shattering resistance associated with domestication is mediated by a NAC gene in soybean. Nat. Commun. 5: 3352.

Funatsuki, H., M. Suzuki, A. Hirose, H. Inaba, T. Yamada, M. Hajika, K. Komatsu, T. Katayama, T. Sayama, M. Ishimoto, and K. Fujino. 2014. Molecular basis of a shattering resistance boosting global dissemination of soybean. Proc. Natl. Acad. Sci. 111(50): 17797–17802.

Gizlice, Z., T.E. Carter, and J.W. Burton. 1994. Genetic base for North American public soybean cultivars released between 1947 and 1988. Crop Sci. 34(5): 1143–1151.

Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire, Q. Sun, and E.S. Buckler. 2014. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. PLoS One 9(2): e90346.

Hodgkin, T., and R. Hajjar. 2008. Using crop wild relatives for crop improvement: Trends and perspectives. p. 535–548. *In* Maxted, N., Ford-Lloyd, B. V., Kell, S.P., Iriondo, J.M., Dulloo, M.E., Turok, J. (eds.), Crop Wild Relative Conservation and Use. CABI, Wallingford, UK.

Hyten, D.L., Q. Song, Y. Zhu, I.Y. 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. 103(45): 16666–16671.

Kim, M., and B.W. Diers. 2013. Fine mapping of the SCN resistance QTL and from PI 468916. Crop Sci. 53(3): 775.

Lam, H.M., X. Xu, X. Liu, W. Chen, G. Yang, F.L. Wong, M.W. Li, W. He, N. Qin, B. Wang, J. Li, M. Jian, J.J. Wang, G. Shao, J.J. Wang, S.S.M. Sun, and G. Zhang. 2010. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. Nat. Genet. 42(12): 1053–1059.

Langmead, B., C. Trapnell, M. Pop, and S.L. Salzberg. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10(3): R25.

Lenth, R. V. 2016. Least-squares means: The R package lsmeans. J. Stat. Softw. 69(1).

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(2): 571–581.

Luo, X., X. Bai, X. Sun, D. Zhu, B. Liu, W. Ji, H. Cai, L. Cao, J. Wu, M. Hu, X. Liu, L. Tang, and Y. Zhu. 2013. Expression of wild soybean WRKY20 in Arabidopsis enhances drought tolerance and regulates ABA signalling. J. Exp. Bot. 64(8): 2155–2169.

Mace, E.S., K.K. Buhariwalla, H.K. Buhariwalla, and J.H. Crouch. 2003. A high-throughput DNA extraction protocol for tropical molecular breeding programs. Plant Mol. Biol. Report. 21(4): 459–460.

Nelson, R.L. 1986. Defining the seed-filling period in soybeans to predict yield. Crop Sci. 26(1): 132–135.

Newell, C.A., X. Delannay, and M.E. Edge. 1987. Interspecific hybrids between the soybean and wild perennial relatives. J. Hered. 78(5): 301–306.

Singh, R.J., and R.L. Nelson. 2015. Intersubgeneric hybridization between *Glycine max* and *G. tomentella*: production of F1, amphidiploid, BC1, BC2, BC3, and fertile soybean plants. Theor. Appl. Genet. 128(6): 1117–1136.

Tang, L., H. Cai, W. Ji, X. Luo, Z. Wang, J. Wu, X. Wang, L. Cui, Y. Wang, Y. Zhu, and X. Bai. 2013. Overexpression of *GsZFP1* enhances salt and drought tolerance in transgenic alfalfa (*Medicago sativa* L.). Plant Physiol. Biochem. 71: 22–30.

Thurber, C.S., J.M. Ma, R.H. Higgins, and P.J. Brown. 2013. Retrospective genomic analysis of sorghum adaptation to temperate-zone grain production. Genome Biol. 14(6): R68.

Yamasaki, M., S.I. Wright, and M.D. McMullen. 2007. Genomic screening for artificial selection during domestication and improvement in maize. Ann. Bot. 100(5): 967–973.

Yang, L., K. Wu, P. Gao, X. Liu, G. Li, and Z. Wu. 2014. GsLRPK, a novel cold-activated leucine-rich repeat receptor-like protein kinase from *Glycine soja*, is a positive regulator to cold stress tolerance. Plant Sci. 215–216: 19–28.

Zhou, Z., Y. Jiang, Z. Wang, Z. Gou, J. Lyu, W. Li, Y. Yu, L. Shu, Y. Zhao, Y. Ma, C. Fang, Y. Shen, T. Liu, C. Li, Q. Li, M. Wu, M. Wang, Y. Wu, Y. Dong, W. Wan, X. Wang, Z. Ding, Y. Gao, H. Xiang, B. Zhu, S.H. Lee, W. Wang, and Z. Tian. 2015. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. Nat. Biotechnol. 33(4): 408–14.