GENETIC MAPPING OF RESISTANCE TO SOYBEAN SUDDEN DEATH SYNDROME AND SOYBEAN OIL QUALITY

By

Paul Joseph Collins

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Plant Breeding, Genetics, and Biotechnology—Crop and Soil Sciences—Doctor of Philosophy

2019

ABSTRACT

GENETIC MAPPING OF RESISTANCE TO SOYBEAN SUDDEN DEATH SYNDROME AND SOYBEAN OIL QUALITY

By

Paul Joseph Collins

Soybean (*Glycine max*) is the world's leading oilseed crop and is a critical source of protein for poultry and swine production. Soybean production is limited by many biotic factors including soybean sudden death syndrome (SDS) which is caused by a soil-borne fungal pathogen, *Fusarium virguliforme*. Effective management methods for soybean sudden death syndrome include long-term rotations, fluopyram seed treatment, and planting SDS resistant varieties. Host resistance to *F. virguliforme* is a quantitative resistance, as it is controlled by many genes, largely of small effect. To more efficiently breed SDS resistant soybean varieties, researchers have sought to identify the loci on the soybean genome responsible for SDS-resistance. Three recombinant inbred line (RIL) populations were evaluated for foliar SDS resistance at a naturally infested field site in Decatur, MI during the 2014 and 2015 growing seasons. These populations segregated for SDS resistance, as they were derived from a parent resistant to SDS and a parent susceptible to SDS. The parents and a subset of RILs from each population were genotyped with the SoySNP6K Illumina Infinium BeadChip. Linkage maps unique to each population were constructed using JoinMap ver. 2. Composite interval mapping was done using WinQTLCartographer (ver. 2.5). Six quantitative trait loci (QTL) were identified to be associated with SDS resistance. Three of the QTL associated with SDS resistance were identified across multiple years and/or populations. While biotic factors, such as SDS, work to limit soybean production, soybean quality factors, such as oil quality, can offer new production opportunities. Soybean oil is predominantly composed of five fatty acids: palmitic acid (11%),

stearic acid (4%), oleic acid (25%), linoleic acid (52%), and linolenic acid (8%). While there is little variability in most commodity soybean varieties for fatty acid content, soybean breeders have been able to introduce oil quality traits into the soybean germplasm. Oil quality traits for soybean oil include high oleic acid content (>75%), low linolenic acid content (<3%), and low saturated fat content (palmitic + stearic < 8%). A RIL population was developed by crossing a high oleic acid, low linolenic acid, and low saturated fat content public variety (E16831) with a high yielding line with a commodity soybean fatty acid profile (E12076T). The parents, the RIL population, and bulked pools representing extreme phenotypes were genotyped with the SoySNP6K Illumina Infinium BeadChip, a linkage map was developed using JoinMap (v.2), and QTLs were detected using WinQTLCartographer (v 2.5). The study confirmed the effects of many known fatty acid biosynthesis genes including *GmFAD2-1B*, *GmFAD3A*, *GmFAD3B*, *GMFAD3C*, and *GmFATB-1A*. The study was also able to dissect possible strategies and limitations of marker assisted selection (MAS) for the development of soybean varieties with high oleic acid, low linolenic acid, and low saturated fat content.

Dedicated to Miss Kimball

ACKNOWLEDGMENTS

I want to offer my sincere gratitude to Dr. Dechun Wang, for his guidance and support throughout graduate school. He has trained me to be a plant breeder and his trust in me has instilled me with the confidence to complete graduate school. I would also like to thank the other members of my graduate guidance committee Dr. Jim Kelly, Dr. Marty Chilvers, and Dr. Eric Olson for their support and accepting positions on my Ph.D. committee. I also want to thank Dr. Dave Douches for serving as the director of the Plant, Breeding, Genetics, and Biotechnology Program and offering me a post-doc position managing the diploid potato breeding program upon graduation.

I also must thank all the members of the MSU Soybean Breeding Lab. Especially fellow students Carmille Bales, Wenyan Du, Shichen Zhang, and Ruijuan Tan. Their guidance in helping me design my experiments, adjust to life as a graduate student, and get the most out of graduate school was irreplaceable. Post-doctoral researchers Dr. Zixiang Wen, Dr. John Yuan, Dr. Umesh Rosyara, Dr. Lin Feng, and Dr. Zhaoming Qi have offered their mentorship and advice at becoming members of the scientific research community, and helped to guide me in designing research to answer questions. Thanks to Dr. Ying-Dong Bi and Dr. Shabir H Wani, visiting scholars, for their friendship and world views. My sincere thanks to technicians John Boyse, Randy Laurenz, and Rob Stoutenburg. Their knowledge of agriculture was critical in my education, and field work would be impossible without their tireless hard work and dedication. I also need to thank Cherry, our greenhouse manager and the hardest worker I have ever had the privilege of working with. She always offered me guidance and valued me as a member of the soybean breeding lab. And a special thanks to Ruijuan and Wen. They helped me in uncountable

ways throughout graduate school and as friends who have been supportive to me throughout all of my graduate school journey.

Lastly, I need to thank my family. My parents have worked incredibly hard to give me the opportunity to achieve all that I have in life. And they instilled the values in me to guide me to where I am today. Their constant support has helped me through many challenges, and I am so grateful for them. My brother, Sean, understands me better than anyone else in the world. Despite being three years older, I've grown to look up to him as an inspiration and role model.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

AN OVERVIEW OF SOYBEAN DOMESTICATION, GENOMIC RESOURCES, PESTS, AND OIL QUALITY

I. Soybean introduction

Soybean (*Glycine max*) is the fourth largest crop in global land area, following wheat, maize, and rice. Since 1960, global soybean production has increased from 17 million metric tons (MMT) to 346 million metric tons in 2018, the fastest rate of increase of any major crop (Hartman et al., 2011). Production is dominated by the U.S., Brazil, and Argentina, which combined to produce 82% of the world's soybeans in 2017 (FAOSTAT, 2017). China is the fourth largest producer, but also is the largest importer, accounting for 64% of global soybean imports in 2016 (FAOSTAT, 2016).

Soybean is an incredibly versatile crop. Direct human consumption includes soy milk, tofu, edamame, natto, miso, and tempeh. However, only 2% of soybeans are consumed directly by humans (Goldsmith, 2008). Composed of approximately 18% oil and 38% protein, the vast majority (98%) of soybean grown today is processed or crushed into protein-rich meal and oil (Ali and Singh, 2010). The meal is used primarily in animal feed, especially in poultry and pork production. Consumption of poultry and pork has increased globally, and soybean production has increased to meet soybean meal demands for poultry and pork production (Goldsmith, 2008). The oil is used in food, as well as in energy and industrial application including cosmetics, plastics, and paint removers.

i. Soybean domestication

Soybean (*Glycine max*) was domesticated approximately 6,000 – 9,000 years ago in China (Carter et al., 2004). A bayesian migration analysis revealed high rates of wild soybean (*G. soja*) introgressions, and high genetic diversity, suggesting that the Huang-Huai Valley, in central China, is the most likely location of soybean domestication (Han et al., 2016). *G. soja* is a wild relative of cultivated soybean and is widely considered to be the progenitor to cultivated soybean. This single origin hypothesis is widely accepted and is supported by microsatellite marker evidence (Guo et al., 2010).

However, synonymous substitution patterns of *G. max* and *G. soja* sequence data suggests that *G. max* and *G. soja* may have diverged as species 0.27 MYA, hundreds of thousands of years before domestication (Kim et al., 2010). The species that split off from *G. soja* 0.27 MYA is referred to as the *G. soja-G. max* complex. In this hypothesis of domestication, the *G. soja-G. max* complex is a hypothetical species that split off from *G. soja* and was the direct progenitor species of *G. max*. A sequencing study with over 500 accessions of *G. soja*, *G. max*, and *G. gracilis* (semi-wild soybeans often classified as landraces or wild soybeans) suggests that *G. gracilis* may be an intermediate transitional species between *G. soja* and *G. max* (Han et al., 2016). It has been suggested that *G. gracilis* may represent the *G. soja–G. max* complex hypothesized to be the progenitor of *G. max* (Sedivy et al., 2017). Analysis of chloroplast genomes of *G. soja* and *G. max* suggests that multiple maternal lines are responsible for soybean domestication, suggesting that soybeans may have been domesticated multiple times in East Asia (Fang et al., 2016). Adding to the complexity of *G. max* domestication, extensive hybridization between *G. max* and *G. soja* occurs with no reproductive barriers between the two

species. Interspecific introgression contributes genetic diversity to *G. max*, but domestication has largely selected against the regions of interspecific introgression (Wang et al., 2019a).

In addition to the bottleneck of domestication, several other bottlenecks have occurred throughout the history of soybean. In North America, bottlenecks occurred during the introduction of landraces from Asia to North America, and during the intensive selective breeding starting in the 1940s. A study of American soybean varieties released from 1947 to 1988 revealed that 99% of the parentage came from 80 landraces (Gizlice et al., 1994). Of those 80 landraces, 17 accounted for 86% of the collective parentage of American soybean varieties for that period. Hyten et al., (2006) examined the effect of these three bottlenecks by examining four soybean populations meant to capture the genetic diversity before, after, and in between these bottlenecks. The four populations were *G. soja* accessions, *G. max* landraces, North American founder lines, and elite varieties released from 1977 to 1990. The analysis determined that diversity loss was largest during the domestication bottleneck. The North American founder lines retained 87% of the genetic diversity of the landraces, but the minor allele frequency was reduced by half. Lastly, the analysis found no significant reduction in nucleotide diversity between North American founder lines and elite varieties. While a large proportion (72%) of the sequence diversity of landraces is present in elite varieties, 79% of the rare alleles present in the landraces are missing in the elite varieties (Hyten et al., 2006). These rare alleles may be critical sources for resistance genes, so landraces and *G. soja* accessions are useful as donors for economically important rare alleles.

While it is important to consider incorporating *G. soja* and *G. max* landraces as sources of useful genetic diversity in elite soybean varieties, it is important to understand the genetic structure within those populations. Resequencing of 302 accessions of cultivated and wild

soybean revealed that phylogenetic groupings correspond with geographic origin (Zhou et al., 2015). Specifically, the authors grouped geographic origin of East Asian soybean accessions from Japan and Korea, Southern China, Northern China, and Northeastern China. The authors found regional differentiation for obvious traits like maturity, but also for less obvious traits such as leaf shape, pubescence color, stem determinacy, and oil content (Zhou et al., 2015). This suggests that *G. soja* accessions and landraces from different regions in East Asia should be considered as separate pools of potential donors of genetic diversity.

ii. Soybean genetic resources

Soybeans are members of Fabaceae, which includes nearly 20,000 species of flowering plants (Doyle et al., 2003). Soybeans belong to the Papilionoideae subfamily, as is true with most crop legumes. Most papilionoid species have a base chromosome number ranging from x=7 to x=11, but soybean has a base chromosome number of x=20. This is the result of an ancient duplication event occurring approximately 13-14 MYA (Schlueter et al., 2004; Schmutz et al., 2010). While cytologically a diploid, this recent whole genome duplication event has classified soybean as a paleopolyploid.

The soybean genome was sequenced using a combination of shotgun sequencing and expressed sequence tags (ESTs) using the variety 'Williams 82' as a reference genome (Schmutz et al., 2010). The reference genome predicted 46,430 protein-coding genes, with nearly 75% of genes present in multiple copies. Since this first assembly (Wm82.a1.v1), a second assembly (Wm82.a2.v1) has been released with improved positions of markers and gene models (Song et al., 2015). The reference genome is accessible through SoyBase, a comprehensive repository for professionally curated genetics, genomics and related data (Grant et al., 2010). SoyBase also

contains curated summaries of quantitative trait loci (QTL) mapping, association mapping, mutant populations, and resequencing studies. In addition to the Williams 82 reference genome, resequencing studies have been conducted for hundreds of additional varieties of *G. soja* and *G. max* (Kim et al., 2010; Zhou et al., 2015; Valliyodan et al., 2016). Two high-throughput single nucleotide polymorphism (SNP) genotyping platforms were developed for soybean: an Illumina Infinium BeadChip with 52,401 SNP markers (SoySNP50K) (Song et al., 2013), and an Affymetrix Axiom with 180,961 SNPs (SoyaSNP180K Axiom) (Lee et al., 2015). Both of these platforms allow for lines to be genotyped with thousands of validated and positioned SNPs in a single well. The design of these chips ensures even genome coverage. These platforms work on a range of genetic material, in fact, the SOYSNP50K was used to genotype the entire U.S. Department of Agriculture (USDA) Soybean Germplasm Collection (Song et al., 2015). This analysis revealed that many lines from the collection were identical or > 99.9% similar. In fact, 4306/18480 (23%) of the *G. max* and 362/1168 (30%) of the *G. soja* accessions were more than 99.9% similar, based on SOYSNP50K data (Song et al., 2015). The collection includes 18,480 *G. max* and 1,168 *G. soja* accessions, and the data from SOYSNP50K is available to public researchers on SoyBase.

II. Challenges for soybean production

Soybean production faces many challenges which can reduce yield and/or grain quality. These factors can be grouped into abiotic and biotic factors. Abiotic factors include the non-living components of the environment, such as water, soil nutrients, salinity, and photoperiod. Drought and flooding often cause large yield losses across the world and in the U.S. Biotic factors include the living components of an environment, such as weeds, insects, nematodes, fungi, bacteria, and

viruses which can affect soybean production. More than 200 pathogens are known to affect soybean, approximately 35 of which are considered to be economically important (Hartman, 2015). Losses due to biotic stresses can have a huge economic impact. As an example, in 2006 the three largest soybean producing countries (U.S., Brazil, and Argentina) lost between 11-13.6% of their yields due to diseases (Hartman, 2015). In addition to selecting for high yield potential, soybean breeders can increase soybean yields by increasing disease resistance, thus reducing the yield losses due to disease.

i. Soybean cyst nematode (SCN)

The highest losses caused by any soybean pest in North America are caused by *Heterodera glycines*, the soybean cyst nematode (SCN). From 2010 to 2014, yield losses due to SCN in the US and Ontario, Canada were estimated to be 617 million bushels, accounting for 28% of the total yield losses due to diseases (Allen et al., 2017). *H. glycines* was first identified in the US in North Carolina in 1954, and is now distributed across all soybean growing regions of North America (Caldwell et al., 1960; Niblack and Riggs, 2015). Juvenile *H. glycines* emerge from eggs, penetrate the soybean roots, and form of a syncytium (feeding site). Feeding of the nematode begins, and the nematodes grow, become sedentary, and develop into adults. Males will exit the feeding site, while females remain attached to the feeding site, growing to the point that they rupture the epidermis of the plant roots. One female can release 50-200 eggs in an external gelatinous matrix. Eggs are also retained within the female's body which develops into a cyst in which the eggs can remain viable in the soil for up to 11 years (Niblack et al., 2006).

a. SCN genetic resistance literature

The first study of the inheritance of SCN resistance was reported in 1960 (Caldwell et al., 1960). This early study used testcrosses and expected segregation ratios to determine that there was at least three independently segregating recessive SCN resistance genes in soybean. These genes were named *rhg1*, *rhg2*, and *rhg3*, 'rhg' being an acronym for 'resistance to *H. glycines*'. Another early study identified a fourth *rhg* gene, a dominant resistance gene linked to a seed coat color locus, naming it *Rhg4* (Matson and Williams, 1965). Advances in genetic marker development, genome sequencing, genetic mapping, and cloning studies have mapped *rhg1* to Chromosome Gm18 and *Rhg4* to Chromosome Gm08 (Cook et al., 2012; Liu et al., 2012).

A review summarized QTL mapping studies published for SCN resistance from 1992 – 2004 (Concibido et al., 2004). This review found 16 papers were published reporting SCN resistance QTL during this period (Weisemann et al., 1992; Concibido et al., 1994; Mahalingam and Skorupska, 1995; Webb et al., 1995; Concibido et al., 1996; Vierling et al., 1996; Chang et al., 1997; Concibido et al., 1997; Heer et al., 1998; Prabhu et al., 1999; Qiu et al., 1999; Meksem et al., 2001; Schuster et al., 2001; Wang et al., 2001; Yue et al., 2001a; Yue et al., 2001b; Webb, 2003). Further investigation has revealed an additional 19 SCN resistance mapping publications since 2004 (Glover et al., 2004; Guo et al., 2005; Kabelka et al., 2005; Ferdous et al., 2006; Guo et al., 2006; Winter et al., 2007; Kazi et al., 2008; Wu et al., 2009; Kazi et al., 2010; Vuong et al., 2010; Chang et al., 2011; Ferreira et al., 2011; Vuong et al., 2011; Arriagada et al., 2012; Abdelmajid et al., 2014; Jiao et al., 2015; Li et al., 2016; Swaminathan et al., 2018). So, at least 35 publications have been reported for QTL mapping of SCN resistance loci, which were found on seventeen of the twenty soybean chromosomes. In addition to QTL studies in cultivated soybean, sources of SCN resistance have been identified in eleven wild, perennial soybean

species (Wen et al., 2017). A stacking study confirmed that QTLs identified from *G. soja* enhanced resistance when stacked with *rhg1* and *Rhg4* (Kim et al., 2011). Despite many studies having identified many other loci responsible for SCN resistance, loci originally described in the 1960's, *rhg1* and *Rhg4* are the most widely studied and most widely deployed in breeding programs today (Concibido et al., 2004; Mitchum, 2016).

Rhg1 was first described in populations derived from the soybean variety 'Peking' (Caldwell et al., 1960). Gene silencing, overexpression, and fluorescence in-situ hybridization revealed that resistance at *Rhg1* was associated with copy number variation at three genes (Glyma.18g022400, Glyma.18g022500 and Glyma.18g022600) on soybean Chromosome Gm18 (Cook et al., 2012). Variation at this locus was assessed over 41 accessions, and it was determined that SCN susceptible varieties contain 1 copy of the above three genes; one group of SCN resistant varieties, which includes 'Peking', have three copies; and another group of resistant varieties, including PI 88788, have seven to ten copies (Cook et al., 2014; Patil et al., 2019). The 'Peking' derived resistant allele at *Rhg1* is referred to as *rhg1-a* or $rhg1_{LC}$, while the PI 88788-derived resistant allele at this locus is referred to as *rhg1-b* or $rhg1_{HC}$. The copy number variation corresponds with previous studies demonstrating that *rhg1-b* is sufficient for SCN resistance, whereas *rhg1-a* requires *Rhg4* to be SCN resistant (Meksem et al., 2001). One of the genes identified, Glyma.18g022500, codes for an *alpha-SNAP* protein, which is a eukaryotically-conserved housekeeping protein involved in vesicle trafficking and hypothesized to be involved in exocytosis-mediated alteration of the SCN feeding site (dit Frey and Robatzek, 2009; Cook et al., 2012). As *alpha-SNAP* proteins serve important cellular housekeeping functions, they are highly conserved across eukaryotes. However, SCN-resistant soybean lines had dysfunctional copies of alpha-SNAP which reduced binding efficiency to another vesicle

trafficking protein, *NSF* (N-ethylmaleimide–sensitive factor) (Bayless et al., 2016). The dysfunctional copies hyperaccumulated around the feeding site, causing targeted cell death, and plant resistance to SCN (Bayless et al., 2016). It is unclear how the cytotoxic *alpha-SNAP* accumulates only in the feeding site and does not cause cell death in the other cells in the plant. However, a novel allele at an *NSF* gene on Chromosome Gm07 (Glyma.07g195900) likely plays a role in mediating the deleterious effects, as it is 100% coinherited with disease resistant *Rhg1* alleles (Bayless et al., 2018).

Based on fine mapping, targeting induced local lesions in genomes (TILLING), virus-induced gene silencing (VIGS), targeted RNA interference (RNAi) it was determined that the *Rhg4* gene was a serine hydroxymethyltransferase (*SHMT*) on Chromosome Gm08 (Glyma.08g108900) (Liu et al., 2012). *SHMT* genes code for enzymes which are responsible for converting serine to glycine. This gene was later functionally validated through transformation, although the resistance phenotype was only partially recovered (Kandoth et al., 2017). 106 soybean lines, including both *G. max* and *G. soja*, were resequenced, phenotyped for SCN resistance, and the sequences at *rhg1* and *Rhg4* were compared (Patil et al., 2019). The resequencing study found three non-synonymous SNP mutations in *GmSHMT08* (Glyma.08g108900) (Patil et al., 2019). Using these SNPs, the accessions were categorized into three haplotype groups, *Rhg4-a* includes the allele from 'Peking', is associated with resistance, and was present in 1 to 4.3 copies; *Rhg4-b* includes the allele from PI 88788, is associated with susceptibility in the absence of *rhg1-b*, and was only present in a single copy; and the *Rhg4-c* allele was exclusively found in PI 436654, was associated with resistance, and was present at 4.3 copies in the genome (Patil et al., 2019). This study also looked at haplotype variation in the

promoter of *GmSHMT08* (Glyma.08g108900) and determined that some promoter haplotypes had moderate increases in resistance across all five races of SCN tested (Patil et al., 2019).

It has been observed that certain isolates of *H. glycines* have overcome the SCN resistance derived from commonly deployed SCN resistance sources (Mitchum, 2016). The lack of genetic diversity of resistance sources used in commercial varieties is certainly a factor in the observed shift in *H. glycines* populations. Over 90% of the SCN resistant varieties available to farmers have resistance derived from PI 88788 (Niblack et al., 2008). The widespread deployment of PI 88788 derived SCN resistance has put pressure on nematode populations to shift in virulence (Niblack et al., 2008). *H. glycines* isolates are categorized based on their response to seven indicator lines (PI 548402, PI 88788, PI 90763, PI 437654, PI 209332, PI 89772, and PI 548316) (Niblack et al., 2003). If the SCN isolate overcomes the resistance of the indicator line, it is given that number it's HG name. For example, isolate 'TN1' is assessed for resistance against all seven indicator lines and the standard susceptible check (Lee 74) in a standard greenhouse assay. Female index (FI) values ([average number of females on the indicator plant / average number of females on Lee 74] * 100) are calculated for all seven indicators against isolate 'TN1'. For indicators 2 (PI 88788), 5 (PI 209332), and 7 (PI 548316), FI > 10. Therefore, 'TN1' is categorized as HG type 2.5.7 because it has overcome the resistance of indicators 2, 5 and 7. While many resistance sources are not resistant against all isolates, some resistance sources have been noted for their broad resistance. Specifically, PI 437654 has been noted for its broad resistance against nearly all isolates of *H. glycines* (Wu et al., 2009).

ii. Soybean sudden death syndrome (SDS)

Soybean sudden death syndrome (SDS) was first reported in Arkansas in 1971 and has since spread to nearly all soybean producing regions of North America. Several species are known to cause SDS, all belonging to soil-borne fungi from clade 2 of the *Fusarium solani* species complex. In South America, F. *virguliforme, F. brasiliense, F. crassistipitatum,* and *F. tucumaniae* are reported to be causal agents of SDS (Aoki et al., 2005; Aoki et al., 2012). In South Africa, *F. virguliforme* and *F. brasiliense* have been reported to be causal agents of SDS (Tewoldemedhin et al., 2014; Tewoldemedhin et al., 2017). In North America *F. virguliforme* had historically been considered the only causal agent of SDS in the U.S., but new evidence includes *F. brasiliense* as another causal agent of SDS in Michigan (Wang et al., 2019b).

Cool, wet springs are the conditions in which yield losses are the most severe. Based on estimates from 2010 to 2014, yield losses due to SDS ranged from 21 – 71 million bushels per year, or $185 - $625 million per year in the U.S. and Canada (Allen et al., 2017). In 2010 and 2014, SDS caused the second and third most yield loss of any soybean disease, respectively (Allen et al., 2017).

F. virguliforme overwinters as chlamydospores, and produces both microconidia and macroconidia (Aoki et al., 2003). There is no known sexual stage for *F. virguliforme* (Hughes et al., 2014). The spores germinate in the soil and form mycelia which infect the roots and can cause symptoms on the soybean roots. These root symptoms include reduced vigor, root discoloration, root rot, and sometimes signs of the fungus which appear as light-blue patches visible on the taproot near the soil line (Hartman et al., 2015). Later in the season, *F. virguliforme* in the root produces toxins which are translocated to the above ground plant tissues (Brar et al., 2011, Chang et al., 2016). Despite the presence of foliar symptoms, *F. virguliforme*

has never been isolated from the above-ground tissues (Roy et al., 1989). Foliar symptoms become most widespread during pod-fill stage (R6 growth stage) of the soybean plant (Fehr and Caviness, 1977). Foliar symptoms include interveinal chlorosis and necrosis. In severe conditions, foliar symptoms can be seen in earlier growth stages, which can result in complete defoliation and up to 100% yield loss (Hartman et al., 2015).

a. SDS genetic resistance literature

From 1996 to the present day, sixteen QTL studies and three genome wide association studies (GWAS) have examined the genomic regions associated with SDS resistance. The QTL studies were mostly conducted using biparental recombinant inbred line (RIL) populations. Resistance sources used for these studies include Forrest (Chang et al., 1996; Hnetkovsky et al., 1996; Chang et al., 1997; Njiti et al, 1998; Iqbal et al., 2001), Pyramid (Njiti et al., 2002), Minsoy (Njiti and Lightfoot, 2006), Ripley (de Farias Neto et al., 2007), PI 567374 (de Farias Neto et al., 2007), Hartwig (Kazi et al., 2008), PI 438489B (Abdelmajid et al., 2012), MD 96-5722 (Anderson et al., 2015) , LS94-3207 (Swaminathan et al., 2016; Swaminathan et al., 2018), LS98-0582(Swaminathan et al., 2016; Swaminathan et al., 2018), MS1606SP (Luckew et al., 2017), LD01-5907 (Tan et al., 2018; Chang et al., 2019), and E07080 (Tan et al., 2019). Phenotyping methods used for QTL mapping studies and GWAS used disease scoring in naturally infested fields (Chang et al., 1996; Hnetkovsky et al., 1996; Chang et al., 1997; Njiti et al, 1998; Iqbal et al., 2001; Njiti et al., 2002; de Farias Neto et al., 2007; Kazi et al., 2008; Wen et al., 2014; Anderson et al., 2015; Tan et al., 2018; Tan et al., 2019), greenhouse assays (Njiti and Lightfoot, 2006; de Farias Neto et al., 2007; Abdelmajid et al., 2012; Zhang et al., 2015; Luckew et al., 2017) and growth chamber assays (Swaminathan et al., 2016; Swaminathan et al.,

2018; Chang et al., 2019). Most of the nineteen QTL mapping or GWAS for SDS resistance used only foliar symptoms to evaluate SDS, but seven of the these studies also included measures of root resistance to SDS (Njiti et al., 1998; Njiti and Lightfoot, 2006; Kazi et al., 2008; Abdelmajid et al., 2012; Zhang et al., 2015; Luckew et al., 2017; Tan et al., 2019). The above QTL mapping and GWAS studies are summarized in Table 1.1. Chang et al. (2018), proposed a novel nomenclature for SDS-resistance loci which improved upon a previous review from 2015 (Lightfoot, 2015). The Chang review used a higher reproducibility standard for SDS resistance loci identified in QTL mapping studies, and proposed an "*Rfv*" nomenclature for ten loci associated with SDS resistance. The standard used required loci to be identified by at least three publications (Chang et al., 2018).

Table 1.1. SDS QTL mapping and GWAS studies*. Studies range from 1996-2018 and include many resistant parents. The environment and phenotyping methods also vary.*

Citation	Year	Resistant Parent	Phenotyping Environment	Root Symptoms Phenotyped
Chang et al., 1996	1996	Forrest	Natural Field	No
Hnetkovsky et al., 1996	1996	Forrest	Natural Field	No
Chang et al., 1997	1997	Forrest	Natural Field	No
Njiti et al., 1998	1998	Forrest	Natural Field	Yes
Iqbal et al., 2001	2001	Forrest	Natural Field	No
Njiti et al., 2002	2002	Pyramid	Natural Field	No
Njiti and Lightfoot, 2006	2006	Minsoy	Greenhouse	Yes
de Farias Neto et al., 2007	2007	Ripley and PI 567374	Greenhouse and Natural Field	No
Kazi et al., 2008	2008	Hartwig	Natural Field	Yes
Abdelmajid et al., 2012	2012	PI 438489B	Greenhouse	Yes
Wen et al., 2014	2014	GWAS	Natural Field	No
Anderson et al., 2015	2015	MD 96-5722	Natural Field	No
Zhang et al., 2015	2015	GWAS	Greenhouse	Yes
Swaminathan et al., 2016	2016	LS94-3207 and LS98-0582	Growth Chamber	No
Luckew et al., 2017	2017	MN1606SP	Greenhouse	Yes
Swaminathan et al., 2018	2018	LS94-3207 and LS98-0582	Growth Chamber	No
Tan et al., 2018	2018	LD01-5907	Natural Field	No
Tan et al., 2019	2019	E07080	Natural Field	Yes
Chang et al., 2019	2019	GWAS + LD01-5907	Growth Chamber	No

b. SDS-SCN connection

F. virguliforme and *H. glycines* often coexist in the same fields, and evidence is mixed as to a synergistic relationship between the two soybean pests (Hartman et al., 2015). A field microplot study looked at the effects of *F. virguliforme*, *H. glycines*, and both pests, finding that the presence of *F. virguliforme* was sufficient to produce foliar symptoms, but with the addition of *H. glycines*, plants showed symptoms earlier, symptoms were more severe, and yield loss was more significant (McLean and Lawrence, 1993). A study examining root tissues of soybean

exposed to *F. virguliforme* and *H. glycines* noted that *F. virguliforme* colonized epidermal and cortical cells of the soybean root adjacent to syncytia formed by *H. glycines* (McLean and Lawrence, 1995). An analysis of high yielding soybean fields in Iowa found that SDS severity increased with increased *F. virguliforme* density, and slightly increased with increased *H. glycines* egg counts (Scherm et al., 1998). A multifactorial microplot experiment revealed that adequate soil moisture was required to observe the synergistic effect that *F. virguliforme* and *H. glycines* that lead to severe SDS foliar and root symptoms (Xing and Westphal, 2013). Another field microplot study incorporated quantitative real-time polymerase chain reaction (qPCR) to detect DNA from *F. virguliforme* into an experiment which demonstrated a synergistic model where *F. virguliforme* was adequate to produce mild SDS symptoms, but required *H. glycines* to produce severe SDS symptoms (Westphal et al., 2014). Most recently, a study which sought to develop models to predict SDS severity and yield loss using risk factors at planting supported a synergistic model where the presence of *H. glycines* in the soil increased SDS severity (Roth et al., 2019).

While many studies support a synergistic relationship between *F. virguliforme* and *H. glycines*, some studies have concluded the opposite. A greenhouse study using inoculations of *F. virguliforme* and *H. glycines* concluded that *H. glycines* did not encourage root colonization by *F. virguliforme* and that *H. glycines* did not increase SDS severity (Gao et al., 2006). A study which analyzed soil samples collected across Wisconsin in 2011 and 2012 determined that *F. virguliforme* and *H. glycines* do not rely on each other for field colonization (Marburger et al., 2014). While the evidence from these two studies seem to contradict evidence collected from other studies, there may be an explanation which accounts for these contradictions. Since soil moisture can affect the ability to detect a synergistic relationship between the two soybean pests

(Xing and Westphal, 2013), it is possible that the studies which concluded there is no synergism between these two pests were conducted in environmental conditions where an environmentally dependent synergism was absent.

As with SDS-SCN synergism, evidence for pleiotropy regarding SDS and SCN resistance genes in unclear. While many studies have mapped genes associated with SDS and SCN resistance, the relationship for host resistance to these two critical pests remains unclear. Some have suggested a model of pleiotropic resistance, citing evidence of synergism between *F. virguliforme* and *H. glycines* with overlapping QTL for resistance to each pest. Other studies have suggested that there is no synergism between the two pests and that resistance genes are coinherited due to tight linkage between independent SDS and SCN resistance genes. While others have argued that pleiotropic resistance, as with synergism, may be dependent upon environmental conditions. The arguments and evidence for these hypotheses of SDS resistance are summarized below.

In a multi-year evaluation of 42 soybean varieties, soybean varieties with resistance to race 6 of *H. glycines* showed resistance to *F. virguliforme*, supporting pleiotropy (Rupe et al., 1991). A gene from the *rhg1/Rfs2* locus was cloned (*GmRLK18-1*, Wm82.a1.v1 Glyma18g02680) and provided nearly complete resistance to *F. virguliforme* foliar and root symptoms, as well as partial resistance to *H. glycines* in greenhouse assays (Srour et al., 2012). These results failed to translate to the field, the gene model used for this conclusion was removed from the Wm82.a2.v1 assembly, and subsequent cloning of Glyma.18g022500 (Wm82.a2.v1) suggests that this gene is more likely to be the causative gene for SCN resistance at this locus (Cook et al., 2012; Bayless et al., 2016). In a study examining the effect of soybean genotypes in multiple locations throughout Kansas, SCN susceptibility increased foliar SDS symptoms

(Brzostowski et al., 2014). In a three-year, multi-state study varieties without SCN resistance had greater SDS foliar symptoms compared with varieties with SCN resistance derived from PI 88788 or 'Peking' (Kandel et al., 2017). Lastly, in a QTL mapping study, two QTLs were detected to correspond with SDS resistance which overlapped with well-known SCN resistance loci *rhg1* and *Rhg4*, suggesting pleiotropy (Tan et al., 2018). The combined action of *rhg1* and *Rhg4* is known to provide SCN resistance for Peking-type resistance (Meksem et al., 2001), and the SDS resistant parent from the SDS resistance QTL mapping study included 'Peking' in its pedigree (Tan et al., 2018).

Other evidence supports the hypotheses that SCN and SDS resistance are often coinherited, not due to pleiotropy, but due to close linkage between independent resistance genes. A study using RFLP and SSR markers to fine map SDS and SCN resistance at the *rhg1/Rfs2* locus in near isogenic lines identified recombinants which support the conclusion that SDS and SCN resistance at this locus comes from independent resistance genes (Meksem et al., 1999). However, SCN resistance at this locus is associated with copy number variation (Cook et al., 2012) which complicates the interpretation of this fine mapping data. Also, an SDS resistance QTL mapping study failed to detect SDS resistance QTL in proximity to *rhg1* despite the presence of polymorphic markers in this region (Kazi et al., 2008). The authors concluded that failing to detect a QTL in this region argues against pleiotropy. Their absence of evidence is falsely interpreted as evidence of absence.

As with evidence supporting and refuting synergism between SDS and SCN, evidence both supports and refutes pleiotropy between SDS and SCN. Variation in environmental conditions may account for the different conclusions present in the literature regarding SDS and SCN pleiotropy. A study examining SDS resistance in a RIL population determined that SDS

resistance was associated with race 3 SCN resistance in some environments, and SDS susceptibility was associated with race 3 and 14 SCN resistance in some environments (Njiti et al., 2002). This evidence, in conjunction with the seemingly contradictory evidence, supports an environmentally dependent pleiotropy model for SDS and SCN. While determining the mechanisms of synergism and pleiotropy of SDS and SCN will be critical towards developing enhanced resistance to both diseases, the best advice for managing both diseases is to use varieties resistant to both diseases (Hartman et al., 2015).

III. Oil quality

Soybean is the largest oilseed crop in the world, accounting for nearly 60% of global oilseed production (Lee et al., 2007). Due to an increase in health-conscious consumers and the FDA ban on trans-fats in 2018, improving soybean oil quality is an area of increasing economic importance. Commodity soybean oil is typically composed of five fatty acids: palmitic (16:0) (11%), stearic (18:0) (4%), oleic (18:1) (25%), linoleic (18:2) (52%), and linolenic (18:3) (8%) (Fehr, 2007). Improving oil quality is based on increasing or decreasing the proportion of the constituent fatty acids, depending on end-use. Increasing the oleic content, while decreasing linoleic and linolenic increases the stability and shelf-life of soybean oil, eliminating the need for chemical hydrogenation, a process which procures trans fats. For industrial applications, increasing linolenic acid can improve the quality of industrial drying oils. Decreasing the saturated fat content (palmitic and stearic) has benefits for cardiovascular health. Increasing saturated fat content increases oxidative stability and shelf life and could be used as a source of trans-fat free margarine or shortening (Fehr, 2007). The Michigan State University Soybean

Breeding Program is developing varieties with low saturated fats (palmitic + stearic < 8%), high oleic acid (> 75%), and low linolenic acid (< 3%).

i. Low saturated fat content

Palmitic and stearic acid are saturated fats which contain 16- and 18- long carbon chains without double bonds, respectively. While other saturated fats are present in soybean oil, they are minor components and rarely measured or selected for when breeding for oil quality. A diet high in saturated fats is associated with poor cardiovascular health due to an increase in low-density lipoprotein cholesterol, so reducing the proportion of saturated fats in soybean oil has benefits to improve cardiovascular health (Li et al., 2015). Foods can be labeled as low in saturated fat if they contain 1 g or less or saturated fats in a 14 g serving, with rounding this is 8.9% saturated fat or less (Fehr, 2007). The goal of soybean breeders is to reduce the saturated fat content of soybean oil from 15% to less than 8%, so that the oil from these varieties can be labeled as low in saturated fat. This has been successful through reducing palmitic acid content from 11% to ~4%.

A reduced palmitic acid mutant was first identified in 1988, when 'Century' was treated with ethyl methanesulfonate (EMS), generating a mutant line, C1726, with a palmitic acid content of 8.6% (Erickson et al., 1988). Another study treated Asgrow 'A1937' with N-nitroso-N-methylurea (NMU) to generate a reduced palmitic acid mutant (A22) with palmitic acid content of 6.8% (Fehr et al., 1991). The progeny of the cross A22 x C1726 showed transgressive segregation for palmitic acid content, with palmitic acid levels as low as 4.4% (Fehr et al., 1991). Later, sequencing of 'Century' and C1726 has revealed that the reduced palmitic acid in C1726 was due to a single SNP in *GmKASIIIA* (Glyma.09g277400) on Chromosome Gm09 which

disrupts donor splice site recognition and increases gene expression of *GmKASIIIA* (Cardinal et al., 2014). The reduced palmitic acid mutant, A22, is a result of a deletion in the *GmFATB-1A* (Glyma.05g012300) gene encoding 16:0-ACP thioesterase activity (Cardinal et al., 2007; De Vries et al., 2011). A naturally occurring mutation at *GmFATB-1A* has also been identified which contributes to the low palmitic phenotype (Burton et al., 1994). Both *GmKASIIIA* and *GmFATB-1A* code for enzymes critical for lipid biosynthesis. The combination of the mutated alleles of *GmKASIIIA* and *GmFATB-1A* has been used to achieve low palmitic (~4%), and thus, low saturated fat (>8%) soybean varieties (Primomo et al., 2002; Gillman et al., 2014).

ii. High oleic acid content

Soybean oil with high levels of oleic acid is a target of many breeding programs, due to the superior resistance to oxidation of the oil without the need for hydrogenation, during which trans-fats are produced as a by-product. High oleic soybean oil production in the U.S. is projected to increase from 140 million pounds in 2016, to 9.3 billion pounds in 2028 (QUALISOY, www.qualisoy.com). To meet the growing demand, Bayer-Monsanto and Corteva have developed soybean varieties with high oleic traits in a strategy which uses a transgenic RNAi approach. Processors across many states pay a $0.50 per bushel premium for high oleic soybeans. Genome editing has also been used to develop a high oleic soybean by a Minnesota biotech company named Calyxt (Haun et al., 2014; Demorest et al., 2016). Their varieties are undergoing large scale quality testing and have had their commercial release in 2018. The USDA has declared that genome-edited crops will not undergo the same regulations transgenic crops currently face. This is promising for innovations in agriculture, but the FDA has yet to announce its policy towards regulation of products derived from genome editing, leaving the future of

genome edited crops uncertain. High oleic soybean varieties have been developed by university breeding programs using mutagenesis breeding which can be sold in markets which prohibit GMO food products, such as organic food companies and export markets where consumer preference demands non-GMO products (European Union, Japan).

FAD2 genes are oil biosynthesis genes which code for fatty acid desaturase enzymes, which convert oleic acid into linoleic acid (Okuley et al., 1994). Therefore, if the *FAD2* genes are absent or nonfunctioning, oleic acid will accumulate in the seed. Silencing of the *GmFAD2* genes in soybean has been conducted using transgenic constructs, genome editing, and mutagenesis, all resulting in the high oleic phenotype. High oleic soybeans were developed by DuPont scientists using a transgenic *GmFAD2-1* gene silencing construct (Kinney, 1995). The high oleic phenotype was also achieved using transgenic gene silencing of mRNA transcripts of *GmFAD2* genes via ribozyme termination (Buhr et al., 2002). Gene editing, using both TALENs (Haun et al., 2014; Demorest et al., 2016) and CRIPSR/Cas9 (Do et al., 2019) gene editing platforms, has been used to target *GmFAD2* genes of soybean to achieve a high oleic phenotype.

Mutagenesis breeding of high oleic soybean varieties requires lines with stacked mutations in *GmFAD2-1A* (Wm82.a2.v1 Glyma.10g278000) and *GmFAD2-1B* ((Wm82.a2.v1 Glyma.20g111000). Mutations in *GmFAD2-1A* are derived from: M23, a line derived from X-ray mutagenesis of 'Bay' (Takagi and Rahman, 1996; Sandhu et al., 2007); KK21, also derived from X-ray mutagenesis of 'Bay' (Anai et al., 2008); 17D, a line derived from a TILLING approach (Dierking and Bilyeu, 2009); PI 603452, which has a naturally occurring frameshift mutation resulting in a premature stop codon (Pham et al., 2011); several unnamed lines from an NMU mutagenized population of 'Williams 82' (Thapa et al., 2016); and a yet to be named line derived from an EMS mutagenesis of '194D', which was itself derived from EMS mutagenesis

of ‘Williams 82’ (Combs and Bilyeu, 2019). Mutants in *GmFAD2-1B* include: PI 283327 which includes a natural missense mutation causing a change in amino acid sequence (Pham et al., 2010); B12, a line derived from an EMS TILLING experiment starting with ‘Bay’ (Hoshino et al., 2010); E11, a line derived from an EMS TILLING experiment starting with ‘Fukuyutaka’ (Hoshino et al., 2010), and a unnamed line derived from an NMU mutagenesis of ‘Williams 82’ (Sweeney et al., 2017). Lines with wildtype alleles at *GmFAD2-1A* and *GmFAD2-1B* have oleic acid contents 20-25%, lines with mutant alleles at either *GmFAD2-1A* or *GmFAD2-1B* have oleic acid contents 30-50%, and lines with both mutant alleles have oleic acid contents > 70% (Alt et al., 2005; Pham et al., 2012; Kulkarni et al., 2017; Sweeney et al., 2017). Therefore, soybean breeders must combine mutant alleles from both *GmFAD2-1A* and *GmFAD2-1B* to achieve varieties which produce high oleic soybean oil.

iii. Low linolenic acid content

As with elevated oleic acid soybean varieties, reduced linolenic acid soybean varieties decrease susceptibility to oxidation, increasing shelf and fry life. This is an alternative to chemical hydrogenation, which produces unhealthy trans-fats. Overexpression of the *FAD3* gene in *Arabidopsis thaliana* has demonstrated that *FAD3* gene products are responsible for the conversion of linoleic acid to linolenic acid (Shah and Xin, 1997). As silencing *GmFAD2* genes increases oleic acid content, silencing *GmFAD3* genes decreases linolenic acid content. Silencing *GmFAD3* genes has been achieved through RNA interference (RNAi) (Flores et al., 2008), and gene editing with TALENs (Demorest et al., 2016). Low linolenic soybean oil has also been achieved through decades of mutagenesis and conventional breeding.

Breeding efforts have been underway by the USDA-ARS to reduce linolenic acid in soybean oil since 1952 (White et al., 1961; Fehr et al., 2007). Mutagenesis and follow up genetic studies have identified three *GmFAD3* genes being responsible for linolenic acid content in soybean seeds, *GmFAD3A* (Wm82.a2.v1 Glyma.14g194300), *GmFAD3B* (Wm82.a2.v1 Glyma.02g227200), and *GmFAD3C* (Wm82.a2.v1 Glyma.18g062000). Mutations in *GmFAD3A* are derived from: A5, a line derived from EMS mutagenesis of FA9525 (derived from PI 80476 and PI 85671) (Hammond and Fehr, 1983; Bilyeu et al., 2003; Pham et al., 2014); C1640, a mutant line derived from EMS mutagenesis of 'Century' (Wilcox et al., 1984; Chappell and Bilyeu, 2006); a natural mutation identified in PI 361088B (Rennie et al., 1988; Chappell and Bilyeu, 2007); a natural mutation identified in PI 123440 (Rennie and Tanner, 1989); M-5 and IL-8, mutant lines developed by X-ray irradiation of 'Bay' (Rahman and Takagi, 1997); J18, a mutant derived from an unpublished X-ray irradiation study, but later characterized and published (Anai et al., 2005); CX1512-44, a line with a complex pedigree which includes mutagenesis (Bilyeu et al., 2005); and PE1690, a line derived from EMS mutagenesis of 'Pungsannamul' (Kim et al., 2015). Mutations in *GmFAD3B* are derived from: A29, a line derived from mutagenesis of A89-144003 (Ross et al., 2000, Bilyeu et al., 2006); and RG10, derived from EMS mutagenesis of the low linolenic acid EMS mutant line C1640 (Stijšin et al., 1998, Reinprecht et al., 2009). Mutations in *GmFAD3C* are derived from: A23, a line derived from EMS mutagenesis of FA47437, originally selected for elevated stearic acid (Bubeck et al., 1989, Bilyeu et al., 2006); and CX1512-44, a line with a complex pedigree which includes mutagenesis, and also includes mutations at FAD3A (Bilyeu et al., 2005). The stacking of *GmFAD3A* mutants with either *GmFAD3B* or *GmFAD3C* is sufficient to generate soybean lines with low linolenic acid (< 3%) (Reinprecht et al., 2009; Bilyeu et al., 2018). The stacking of

mutant alleles at *GmFAD3A*, *GmFAD3B*, and *GmFAD3C* generates soybean lines with ultra-low levels of linolenic acid (< 1%) (Ross et al., 2000; Bilyeu et al., 2006; Bilyeu et al., 2011). The low linolenic acid content trait can be combined with the high oleic acid content trait (Pham et al., 2012; Bilyeu et al., 2018).

IV. Aims of the dissertation research

The research presented in this dissertation intends to identify associations between the soybean genome, and resistance to sudden death syndrome, low saturated fat content, high oleic acid content, and low linolenic acid content. As noted in the literature review, these traits been extensively studied to identify genes associated with these traits.

In Chapter 2, a multi-year QTL study was conducted in a field naturally infested with SDS. This study included three recombinant inbred line populations designed to map SDS resistance QTL. These populations have SDS resistance derived from two MSU breeding lines, E09088 and E09014. As the literature review noted, SCN may have a synergistic effect on SDS symptom development, and SCN was present in the fields used for the QTL study. In addition to being SDS resistant, E09088 and E09014 both have SCN resistance derived from PI 88788. Therefore, SCN literature was also summarized above, as it likely played a role in the SDS QTL mapping study. The study in this chapter also included spatial data adjustment, which was applied in an *ad hoc* manner due to uneven pathogen pressure in the field where the study was conducted.

In Chapter 3, a study was conducted on a recombinant inbred line population which segregated for the low saturated fat trait, the high oleic acid trait, and the low linolenic acid trait. As noted in Chapter 1, most of these traits were introduced into the soybean germplasm by

mutagenesis, and the causative genes for these traits are well characterized. While studies have examined populations with combinations of two of these three traits, few studies have examined the genetic associations in a population segregating for all three oil quality traits. The aim of this chapter was to characterize which genes from the literature were present in MSU breeding lines, as well as identify other genes critical towards achieving oil quality targets.

LITERATURE CITED

LITERATURE CITED

Abdelmajid, K. M., Ramos, L., Leandro, L., Mbofung, G., Hyten, D. L., Kantartzi, S. K., et al. (2012). The 'PI 438489B' by 'Hamilton' SNP-based genetic linkage map of soybean [*Glycine max* (L.) Merr.] identified quantitative trait loci that underlie seedling SDS resistance. Journal of Plant Genome Sciences 1, 18-30.

Abdelmajid, K. M., Ramos, L., Hyten, D. L., Bond, J., Bendahmane, A., Arelli, P. R., et al. (2014). Quantitative trait loci (QTL) that underlie SCN resistance in soybean [*Glycine max* (L.) Merr.] PI438489B by 'Hamilton' recombinant inbred line (RIL) population. Atlas Journal of Plant Biology 3, 29-38.

Ali, N., and Singh, G. (2010). Chapter 16: Soybean processing and utilization. in The soybean: botany, production and uses, eds Guriqbal Singh (CAB International Publishing), 345-374.

Alt, J. L., Fehr, W. R., Welke, G. A., and Shannon, J. G. (2005). Transgressive segregation for oleate content in three soybean populations. Crop Science, 45, 2005-2007.

Anai, T., Yamada, T., Kinoshita, T., Rahman, S. M., and Takagi, Y. (2005). Identification of corresponding genes for three low-α-linolenic acid mutants and elucidation of their contribution to fatty acid biosynthesis in soybean seed. Plant Science, 168, 1615-1623.

Anai, T., Yamada, T., Hideshima, R., Kinoshita, T., Rahman, S. M., et al. (2008). Two high-oleic-acid soybean mutants, M23 and KK21, have disrupted microsomal omega-6 fatty acid desaturase, encoded by *GmFAD2-1a*. Breeding Science, 58, 447-452.

Anderson, J., Akond, M., Kassem, M. A., Meksem, K., and Kantartzi, S. K. (2015). Quantitative trait loci underlying resistance to sudden death syndrome (SDS) in MD96-5722 by 'Spencer' recombinant inbred line population of soybean. Biotech, 5, 203-210.

Aoki, T., O'Donnell, K., Homma, Y., and Lattanzi, A. R. (2003). Sudden-death syndrome of soybean is caused by two morphologically and phylogenetically distinct species within the *Fusarium solani* species complex—*F. virguliforme* in North America and *F. tucumaniae* in South America. Mycologia, 95, 660-684.

Aoki, T., O'Donnell, K., and Scandiani, M. M. (2005). Sudden death syndrome of soybean in South America is caused by four species of Fusarium: *Fusarium brasiliense* sp. nov., *F. cuneirostrum* sp. nov., *F. tucumaniae*, and *F. virguliforme*. Mycoscience, 46, 162-183.

Aoki, T., Scandiani, M. M., and O'Donnell, K. (2012). Phenotypic, molecular phylogenetic, and pathogenetic characterization of *Fusarium crassistipitatum* sp. nov., a novel soybean sudden death syndrome pathogen from Argentina and Brazil. Mycoscience, 53, 167-186.

Arriagada, O., Mora, F., Dellarossa, J. C., Ferreira, M. F., Cervigni, G. D., et al. (2012). Bayesian mapping of quantitative trait loci (QTL) controlling soybean cyst nematode resistant. Euphytica, 186, 907-917.

Bayless, A. M., Smith, J. M., Song, J., McMinn, P. H., Teillet, A., et al. (2016). Disease resistance through impairment of α-SNAP–NSF interaction and vesicular trafficking by soybean *Rhg1*. Proceedings of the National Academy of Sciences, 113, E7375-E7382.

Bayless, A. M., Zapotocny, R. W., Grunwald, D. J., Amundson, K. K., Diers, B. W., et al. (2018). An atypical N-ethylmaleimide sensitive factor enables the viability of nematode-resistant *Rhg1* soybeans. Proceedings of the National Academy of Sciences, 115, E4512-E4521.

Bilyeu, K. D., Palavalli, L., Sleper, D. A., and Beuselinck, P. R. (2003). Three microsomal omega-3 fatty-acid desaturase genes contribute to soybean linolenic acid levels. Crop Science, 43, 1833-1838.

Bilyeu, K., Palavalli, L., Sleper, D., and Beuselinck, P. (2005). Mutations in soybean microsomal omega-3 fatty acid desaturase genes reduce linolenic acid concentration in soybean seeds. Crop Science, 45, 1830-1836.

Bilyeu, K., Palavalli, L., Sleper, D. A., and Beuselinck, P. (2006). Molecular genetic resources for development of 1% linolenic acid soybeans. Crop Science, 46, 1913-1918.

Bilyeu, K., Gillman, J. D., and LeRoy, A. R. (2011). Novel FAD3 mutant allele combinations produce soybeans containing 1% linolenic acid in the seed oil. Crop science, 51(1), 259-264.

Bilyeu, K., Škrabišová, M., Allen, D., Rajcan, I., Palmquist, D. E., Gillen, A., et al. (2018). The interaction of the soybean seed high oleic acid oil trait with other fatty acid modifications. Journal of the American Oil Chemists' Society, 95, 39-49.

Brar, H. K., Swaminathan, S., and Bhattacharyya, M. K. (2011). The *Fusarium virguliforme* toxin FvTox1 causes foliar sudden death syndrome-like symptoms in soybean. Molecular Plant-Microbe Interactions, 24, 1179-1188.

Brzostowski, L. F., Schapaugh, W. T., Rzodkiewicz, P. A., Todd, T. C., and Little, C. R. (2014). Effect of host resistance *to Fusarium virguliforme* and *Heterodera glycines* on sudden death syndrome disease severity and soybean yield. Plant Health Progress, 15, 1-8.

Bubeck, D. M., Fehr, W. R., and Hammond, E. G. (1989). Inheritance of palmitic and stearic acid mutants of soybean. Crop Science, 29, 652-656.

Buhr, T., Sato, S., Ebrahim, F., Xing, A., Zhou, Y., et al. (2002). Ribozyme termination of RNA transcripts down-regulate seed fatty acid genes in transgenic soybean. The Plant Journal, 30, 155-163.

Burton, J. W., Wilson, R. F., and Brim, C. A. (1994). Registration of N79-2077-12 and N87-2122-4, two soybean germplasm lines with reduced palmitic acid in seed oil. Crop Science, 34, 313.

Caldwell, B. E., Brim, C. A., and Ross, J. P. (1960). Inheritance of resistance of soybeans to the cyst nematode, *Heterodera glycines*, Agronomy Journal, 52, 635-636.

Cardinal, A. J., Burton, J. W., Camacho-Roger, A. M., Yang, J. H., Wilson, R. F., et al. (2007). Molecular analysis of soybean lines with low palmitic acid content in the seed oil. Crop Science, 47, 304-310.

Cardinal, A. J., Whetten, R., Wang, S., Auclair, J., Hyten, D., et al. (2014). Mapping the low palmitate *fap1* mutation and validation of its effects in soybean oil and agronomic traits in three soybean populations. Theoretical and Applied Genetics, 127, 97-111.

Carter, T. E., Nelson, R. L., Sneller, C. H., Cui, Z., Boerma, H. R., et al. (2004). 3. Genetic diversity in soybean. in Soybeans: Improvement, production, and uses, eds H. R. Boerma and J. E. Specht (Am. Soc. of Agron.), 114.

Chang, S. J. C., Doubler, T. W., Kilo, V., Suttner, R., Klein, J., et al. (1996). Two additional loci underlying durable field resistance to soybean sudden death syndrome (SDS). Crop Science, 36, 1684-1688.

Chang, S. J. C., Doubler, T. W., Kilo, V. Y., Abu-Thredeih, J., Prabhu, R., et al. (1997). Association of loci underlying field resistance to soybean sudden death syndrome (SDS) and cyst nematode (SCN) race 3. Crop Science, 37, 965-971.

Chang, W., Dong, L., Wang, Z., Hu, H., Han, Y., et al. (2011). QTL underlying resistance to two HG types of *Heterodera glycines* found in soybean cultivar'L-10'. BMC Genomics, 12, 233.

Chang, H. X., Domier, L. L., Radwan, O., Yendrek, C. R., Hudson, M. E., and Hartman, G. L. (2016). Identification of multiple phytotoxins produced by *Fusarium virguliforme* including a phytotoxic effector (FvNIS1) associated with sudden death syndrome foliar symptoms. Molecular Plant-Microbe Interactions, 29, 96-108.

Chang, H. X., Roth, M. G., Wang, D., Cianzio, S. R., Lightfoot, D. A., Hartman, G. L., et al. (2018). Integration of sudden death syndrome resistance loci in the soybean genome. Theoretical and Applied Genetics, 131, 757-773.

Chang, H. X., Tan, R., Hartman, G. L., Wen, Z., Sang, H., Domier, L. L., et al. (2019). Characterization of soybean STAY-GREEN genes in susceptibility to foliar chlorosis of sudden death syndrome. Plant Physiology, 180, 711-717.

Chappell, A. S., and Bilyeu, K. D. (2006). A *GmFAD3A* mutation in the low linolenic acid soybean mutant C1640. Plant Breeding, 125, 535-536.

Chappell, A. S., and K. D. Bilyeu. (2007). The low linolenic acid soybean line PI 361088B contains a novel *GmFAD3A* mutation. Crop Science, 47, 1705-1710.

Combs, R., and Bilyeu, K. (2019). Novel alleles of *FAD2-1A* induce high levels of oleic acid in soybean oil. Molecular Breeding, 39, 79.

Cook, D. E., Lee, T. G., Guo, X., Melito, S., Wang, K., Bayless, A. M., et al. (2012). Copy number variation of multiple genes at *Rhg1* mediates nematode resistance in soybean. Science, 338, 1206-1209.

Cook, D. E., Bayless, A. M., Wang, K., Guo, X., Song, Q., Jiang, J., et al. (2014). Distinct copy number, coding sequence, and locus methylation patterns underlie *Rhg1*-mediated soybean resistance to soybean cyst nematode. Plant Physiology, 165, 630-647.

Concibido, V. C., Denny, R. L., Boutin, S. R., Hautea, R., Orf, J. H., and Young, N. D. (1994). DNA marker analysis of loci underlying resistance to soybean cyst nematode (*Heterodera glycines* Ichinohe). Crop Science, 34, 240-246.

Concibido, V. C., Young, N. D., Lange, D. A., Denny, R. L., Danesh, D., and Orf, J. H. (1996). Targeted comparative genome analysis and qualitative mapping of a major partial-resistance gene to the soybean cyst nematode. Theoretical and Applied Genetics, 93, 234-241.

Concibido, V. C., Lange, D. A., Denny, R. L., Orf, J. H., and Young, N. D. (1997). Genome mapping of soybean cyst nematode resistance genes in 'Peking', PI 90763, and PI 88788 using DNA markers. Crop Science, 37, 258-264.

Concibido, V. C., Diers, B. W., and Arelli, P. R. (2004). A decade of QTL mapping for cyst nematode resistance in soybean. Crop Science, 44, 1121-1131.

de Farias Neto, A. L., Hashmi, R., Schmidt, M., Carlson, S. R., Hartman, G. L., Li, S., et al. (2007). Mapping and confirmation of a new sudden death syndrome resistance QTL on linkage group D2 from the soybean genotypes PI 567374 and 'Ripley'. Molecular Breeding, 20, 53-62.

De Vries, B. D., Fehr, W. R., Welke, G. A., and Dewey, R. E. (2011). Molecular characterization of the mutant *fap3* (A22) allele for reduced palmitate concentration in soybean. Crop Science, 51, 1611-1616.

Demorest, Z. L., Coffman, A., Baltes, N. J., Stoddard, T. J., Clasen, B. M., Luo, S., et al. (2016). Direct stacking of sequence-specific nuclease-induced mutations to produce high oleic and low linolenic soybean oil. BMC Plant Biology, 16, 225.

Dierking, E. C., and Bilyeu, K. D. (2009). New sources of soybean seed meal and oil composition traits identified through TILLING. BMC Plant Biology, 9, 89.

dit Frey, N. F., and Robatzek, S. (2009). Trafficking vesicles: pro or contra pathogens?. Current Opinion in Plant Biology, 12, 437-443.

Do, P. T., Nguyen, C. X., Bui, H. T., Tran, L. T., Stacey, G., Gillman, J. D., et al. (2019). Demonstration of highly efficient dual gRNA CRISPR/Cas9 editing of the homeologous *GmFAD2–1A* and *GmFAD2–1B* genes to yield a high oleic, low linoleic and α-linolenic acid phenotype in soybean. BMC Plant Biology, 19, 311.

Doyle, J. J., and Luckow, M. A. (2003). The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. Plant Physiology, 131, 900-910.

Fang, C., Ma, Y., Yuan, L., Wang, Z., Yang, R., Zhou, Z., ... and Tian, Z. (2016). Chloroplast DNA underwent independent selection from nuclear genes during soybean domestication and improvement. Journal of Genetics and Genomics, 43, 217.

FAOSTAT. (2016). FAOSTAT database collections. Food and Agriculture Organization of the United Nations. Available at: http://www.fao.org/faostat/ [Accessed August 27, 2019].

FAOSTAT. (2017). FAOSTAT database collections. Food and Agriculture Organization of the United Nations. Available at: http://www.fao.org/faostat/ [Accessed August 27, 2019].

Fehr, W. R., and Caviness, C. E. (1977). Stages of soybean development. Special Report, 87, 2-12.

Fehr, W. R., Welke, G. A., Cianzio, S. R., Duvick, D. N., and Hammond, E. G. (1991). Inheritance of reduced palmitic acid content in seed oil of soybean. Crop Science, 31, 88-89.

Fehr, W. R. (2007). Breeding for modified fatty acid composition in soybean. Crop Science, 47(Supplement_3), S-72.

Ferdous, S. A., Watanabe, S., Suzuki-Orihara, C., Tanaka, Y., Kamiya, M., Yamanaka, N., et al. (2006). QTL analysis of resistance to soybean cyst nematode race 3 in soybean cultivar Toyomusume. Breeding Science, 56, 155-163.

Ferreira, M. F. D. S., Cervigni, G. D. L., Ferreira, A., Schuster, I., Santana, F. A., Pereira, W. D., et al. (2011). QTLs for resistance to soybean cyst nematode, races 3, 9, and 14 in cultivar Hartwig. Pesquisa Agropecuária Brasileira, 46, 420-428.

Flores, T., Karpova, O., Su, X., Zeng, P., Bilyeu, K., Sleper, D. A., et al. (2008). Silencing of *GmFAD3* gene by siRNA leads to low α-linolenic acids (18: 3) of *fad3*-mutant phenotype in soybean [*Glycine max* (Merr.)]. Transgenic Research, 17, 839-850.

Gao, X., Jackson, T. A., Hartman, G. L., and Niblack, T. L. (2006). Interactions between the soybean cyst nematode and *Fusarium solani* f. sp. *glycines* based on greenhouse factorial experiments. Phytopathology, 96, 1409-1415.

Gelin, J. R., Arelli, P. R., and Rojas-Cifuentes, G. A. (2006). Using independent culling to screen plant introductions for combined resistance to soybean cyst nematode and sudden death syndrome. Crop Science, 46, 2081-2083.

Gillman, J. D., Tetlow, A., Hagely, K., Boersma, J. G., Cardinal, A., Rajcan, I., et al. (2014). Identification of the molecular genetic basis of the low palmitic acid seed oil trait in soybean mutant line RG3 and association analysis of molecular markers with elevated seed stearic acid and reduced seed palmitic acid. Molecular Breeding, 34, 447-455.

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

Glover, K. D., Wang, D., Arelli, P. R., Carlson, S. R., Cianzio, S. R., and Diers, B. W. (2004). Near isogenic lines confirm a soybean cyst nematode resistance gene from PI 88788 on linkage group J. Crop Science, 44, 936-941.

Guo, B., Sleper, D. A., Arelli, P. R., Shannon, J. G., and Nguyen, H. T. (2005). Identification of QTLs associated with resistance to soybean cyst nematode races 2, 3 and 5 in soybean PI 90763. Theoretical and Applied Genetics, 111, 965-971.

Guo, B., Sleper, D. A., Nguyen, H. T., Arelli, P. R., and Shannon, J. G. (2006). Quantitative trait loci underlying resistance to three soybean cyst nematode populations in soybean PI 404198A. Crop Science, 46, 224-233.

Guo, J., Wang, Y., Song, C., Zhou, J., Qiu, L., Huang, H., et al. (2010). A single origin and moderate bottleneck during domestication of soybean (*Glycine max*): implications from microsatellites and nucleotide sequences. Annals of Botany, 106, 505-514.

Han, Y., Zhao, X., Liu, D., Li, Y., Lightfoot, D. A., Yang, Z., et al. (2016). Domestication footprints anchor genomic regions of agronomic importance in soybeans. New Phytologist, 209, 871-884.

Hartman, G. L., West, E. D., and Herman, T. K. (2011). Crops that feed the World 2. Soybean—worldwide production, use, and constraints caused by pathogens and pests. Food Security, 3, 5-17.

Hartman, G. L. (2015). Worldwide importance of soybean pathogens and pests. Compendium of Soybean Diseases and Pests, 4-5.

Hartman, G. L., Chang, H. X., and Leandro, L. F. (2015). Research advances and management of soybean sudden death syndrome. Crop Protection, 73, 60-66.

Haun, W., Coffman, A., Clasen, B. M., Demorest, Z. L., Lowy, A., Ray, E., et al. (2014). Improved soybean oil quality by targeted mutagenesis of the fatty acid desaturase 2 gene family. Plant Biotechnology Journal, 12, 934-940.

Heer, J. A., Knap, H. T., Mahalingam, R., Shipe, E. R., Arelli, P. R., and Matthews, B. F. (1998). Molecular markers for resistance to *Heterodera glycines* in advanced soybean germplasm. Molecular Breeding, 4, 359-367.

Hnetkovsky, N., Chang, S. J. C., Doubler, T. W., Gibson, P. T., and Lightfoot, D. A. (1996). Genetic mapping of loci underlying field resistance to soybean sudden death syndrome (SDS). Crop Science, 36, 393-400.

Hoshino, T., Takagi, Y., and Anai, T. (2010). Novel *GmFAD2-1b* mutant alleles created by reverse genetics induce marked elevation of oleic acid content in soybean seeds in combination with *GmFAD2-1a* mutant alleles. Breeding Science, 60, 419-425.

Hughes, T. J., O'Donnell, K., Sink, S., Rooney, A. P., Scandiani, M. M., Luque, A., et al. (2014). Genetic architecture and evolution of the mating type locus in fusaria that cause soybean sudden death syndrome and bean root rot. Mycologia, 106, 686-697.

Hyten, D. L., Song, Q., Zhu, Y., Choi, I. Y., Nelson, R. L., Costa, J. M., et al. (2006). Impacts of genetic bottlenecks on soybean genome diversity. Proceedings of the National Academy of Sciences, 103, 16666-16671.

Iqbal, M. J., Meksem, K., Njiti, V. N., Kassem, M. A., and Lightfoot, D. A. (2001). Microsatellite markers identify three additional quantitative trait loci for resistance to soybean sudden-death syndrome (SDS) in Essex× Forrest RILs. Theoretical and Applied Genetics, 102, 187-192.

Jiao, Y., Vuong, T. D., Liu, Y., Meinhardt, C., Liu, Y., Joshi, T., et al. (2015). Identification and evaluation of quantitative trait loci underlying resistance to multiple HG types of soybean cyst nematode in soybean PI 437655. Theoretical and Applied Genetics, 128, 15-23.

Kabelka, E. A., Carlson, S. R., and Diers, B. W. (2005). Localization of two loci that confer resistance to soybean cyst nematode from *Glycine soja* PI 468916. Crop Science, 45, 2473-2481.

Kandel, Y. R., Wise, K. A., Bradley, C. A., Chilvers, M. I., Byrne, A. M., Tenuta, A. U., et al. (2017). Effect of soybean cyst nematode resistance source and seed treatment on population densities of *Heterodera glycines*, sudden death syndrome, and yield of soybean. Plant Disease, 101, 2137-2143.

Kandoth, P. K., Liu, S., Prenger, E., Ludwig, A., Lakhssassi, N., Heinz, R., et al. (2017). Systematic mutagenesis of serine hydroxymethyltransferase reveals an essential role in nematode resistance. Plant Physiology, 175, 1370-1380.

Kazi, S., Shultz, J., Afzal, J., Johnson, J., Njiti, V. N., and Lightfoot, D. A. (2008). Separate loci underlie resistance to root infection and leaf scorch during soybean sudden death syndrome. Theoretical and Applied Genetics, 116, 967-977.

Kazi, S., Shultz, J., Afzal, J., Hashmi, R., Jasim, M., Bond, J., et al. (2010). Iso-lines and inbred-lines confirmed loci that underlie resistance from cultivar 'Hartwig' to three soybean cyst nematode populations. Theoretical and Applied Genetics, 120, 633-644.

Kim, M. Y., Lee, S., Van, K., Kim, T. H., Jeong, S. C., Choi, I. Y., et al. (2010). Whole-genome sequencing and intensive analysis of the undomesticated soybean (*Glycine soja* Sieb. and Zucc.) genome. Proceedings of the National Academy of Sciences, 107, 22032-22037.

Kim, M., Hyten, D. L., Niblack, T. L., and Diers, B. W. (2011). Stacking resistance alleles from wild and domestic soybean sources improves soybean cyst nematode resistance. Crop Science, 51, 934-943.

Kim, M., Song, J. T., Bilyeu, K. D., and Lee, J. D. (2015). A new low linolenic acid allele of *GmFAD3A* gene in soybean PE1690. Molecular Breeding, 35, 155.

Kinney, A. J. (1995). Improving soybean seed quality. in Induced mutations and molecular techniques for crop improvement. (International Atomic Energy Agency), 101-113.

Kulkarni, K. P., Kim, M., Song, J. T., Bilyeu, K. D., and Lee, J. D. (2017). Genetic improvement of the fatty acid biosynthesis system to alter the ω-6/ω-3 ratio in the soybean seed. Journal of the American Oil Chemists' Society, 94, 1403-1410.

Lee, J. D., Bilyeu, K. D., and Shannon, J. G. (2007). Genetics and breeding for modified fatty acid profile in soybean seed oil. Journal Crop Sci. Biotech., 10, 201-210.

Lee, Y. G., Jeong, N., Kim, J. H., Lee, K., Kim, K. H., Pirani, A., et al. (2015). Development, validation and genetic analysis of a large soybean SNP genotyping array. The Plant Journal, 81, 625-636.

Li, Y., Hruby, A., Bernstein, A. M., Ley, S. H., Wang, D. D., Chiuve, S. E., et al. (2015). Saturated fats compared with unsaturated fats and sources of carbohydrates in relation to risk of coronary heart disease: a prospective cohort study. Journal of the American College of Cardiology, 66, 1538-1548.

Li, Y. H., Shi, X. H., Li, H. H., Reif, J. C., Wang, J. J., Liu, Z. X., et al. (2016). Dissecting the genetic basis of resistance to soybean cyst nematode combining linkage and association mapping. The Plant Genome, 9.

Lightfoot, D. A. (2015). Two decades of molecular marker-assisted breeding for resistance to soybean sudden death syndrome. Crop Science, 55, 1460-1484.

Liu, S., Kandoth, P. K., Warren, S. D., Yeckel, G., Heinz, R., Alden, J., et al. (2012). A soybean cyst nematode resistance gene points to a new mechanism of plant resistance to pathogens. Nature, 492, 256-260.

Luckew, A. S., Swaminathan, S., Leandro, L. F., Orf, J. H., and Cianzio, S. R. (2017). 'MN1606SP'by 'Spencer' filial soybean population reveals novel quantitative trait loci and interactions among loci conditioning SDS resistance. Theoretical and Applied Genetics, 130, 2139-2149.

Mahalingam, R., and Skorupska, H. T. (1996). Cytological expression of early response to infection by *Heterodera glycines* Ichinohe in resistant PI 437654 soybean. Genome, 39, 986-998.

Marburger, D., Conley, S., Esker, P., MacGuidwin, A., and Smith, D. (2014). Relationship between *Fusarium virguliforme* and *Heterodera glycines* in commercial soybean fields in Wisconsin. Plant Health Progress, 15, 11-18.

McLean, K. S., and Lawrence, G. W. (1993). Interrelationship of *Heterodera glycines* and *Fusarium solani* in sudden death syndrome of soybean. Journal of Nematology, 25, 434-439.

McLean, K. S., and Lawrence, G. W. (1995). Development of *Heterodera glycines* as affected by *Fusarium solani*, the causal agent of sudden death syndrome of soybean. Journal of Nematology, 27, 70-77.

Meksem, K., Doubler, T. W., Chancharoenchai, K., Nijti, N., Chang, S. J., Arelli, A. R., et al. (1999). Clustering among loci underlying soybean resistance to *Fusarium solani*, SDS and SCN in near-isogenic lines. Theoretical and Applied Genetics, 99, 1131-1142.

Meksem, K., Pantazopoulos, P., Njiti, V. N., Hyten, L. D., Arelli, P. R., and Lightfoot, D. A. (2001). 'Forrest' resistance to the soybean cyst nematode is bigenic: saturation mapping of the *Rhg1*and *Rhg4* loci. Theoretical and Applied Genetics, 103, 710-717.

Mitchum, M. G. (2016). Soybean resistance to the soybean cyst nematode *Heterodera glycines*: an update. Phytopathology, 106, 1444-1450.

Niblack, T. L., Noel, G. R., and Lambert, K. L. (2003). The Illinois SCN type test: practical application of the Hg type classification system. J. Nematol., 35, 345-355.

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

Niblack, T. L., Colgrove, A. L., Colgrove, K., and Bond, J. P. (2008). Shift in virulence of soybean cyst nematode is associated with use of resistance from PI 88788. Plant Health Progress, 9, 29.

Niblack, T. L., and Riggs, R. D. (2015) Soybean cyst nematode. in Compendium of soybean diseases. (APS Press), 52-53.

Njiti, V. N., Doubler, T. W., Suttner, R. J., Gray, L. E., Gibson, P. T., and Lightfoot, D. A. (1998). Resistance to soybean sudden death syndrome and root colonization by *Fusarium solani* f. sp. glycine in near-isogenic lines. Crop Science, 38, 472-477.

Njiti, V. N., Meksem, K., Iqbal, M. J., Johnson, J. E., Kassem, M. A., Zobrist, K. F., et al. (2002). Common loci underlie field resistance to soybean sudden death syndrome in Forrest, Pyramid, Essex, and Douglas. Theoretical and Applied Genetics, 104, 294-300.

Njiti, V. N., and Lightfoot, D. A. (2006). Genetic analysis infers Dt loci underlie resistance to *Fusarium solani* f. sp. glycines in indeterminate soybeans. Canadian Journal of Plant Science, 86, 83-90.

Okuley, J., Lightner, J., Feldmann, K., Yadav, N., and Lark, E. (1994). Arabidopsis *FAD2* gene encodes the enzyme that is essential for polyunsaturated lipid synthesis. The Plant Cell, 6, 147-158.

Patil, G. B., Lakhssassi, N., Wan, J., Song, L., Zhou, Z., Klepadlo, M., et al. (2019). Whole-genome re-sequencing reveals the impact of the interaction of copy number variants of the *rhg1* and *Rhg4* genes on broad-based resistance to soybean cyst nematode. Plant Biotechnology Journal, 17, 1595–1611.

Pham, A. T., Lee, J. D., Shannon, J. G., and Bilyeu, K. D. (2010). Mutant alleles of *FAD2-1A* and *FAD2-1B* combine to produce soybeans with the high oleic acid seed oil trait. BMC Plant Biology, 10, 195.

Pham, A. T., Lee, J. D., Shannon, J. G., and Bilyeu, K. D. (2011). A novel *FAD2-1A* allele in a soybean plant introduction offers an alternate means to produce soybean seed oil with 85% oleic acid content. Theoretical and Applied Genetics, 123, 793-802.

Pham, A. T., Shannon, J. G., and Bilyeu, K. D. (2012). Combinations of mutant *FAD2* and *FAD3* genes to produce high oleic acid and low linolenic acid soybean oil. Theoretical and Applied Genetics, 125, 503-515.

Pham, A. T., Bilyeu, K., Chen, P., Boerma, H. R., and Li, Z. (2014). Characterization of the *fan1* locus in soybean line A5 and development of molecular assays for high-throughput genotyping of *FAD3* genes. Molecular Breeding, 33, 895-907.

Prabhu, R. R., Njiti, V. N., Bell-Johnson, B., Johnson, J. E., Schmidt, M. E., Klein, J. H., et al. (1999). Selecting soybean cultivars for dual resistance to soybean cyst nematode and sudden death syndrome using two DNA markers. Crop Science, 39, 982-987.

Primomo, V. S., Falk, D. E., Ablett, G. R., Tanner, J. W., and Rajcan, I. (2002). Inheritance and interaction of low palmitic and low linolenic soybean. Crop Science, 42, 31-36.

Qiu, B. X., Arelli, P. R., and Sleper, D. A. (1999). RFLP markers associated with soybean cyst nematode resistance and seed composition in a 'Peking' × 'Essex' population. Theoretical and Applied Genetics, 98, 356-364.

Rahman, S. M., and Takagi, Y. (1997). Inheritance of reduced linolenic acid content in soybean seed oil. Theoretical and Applied Genetics, 94, 299-302.

Roth, M. G., Noel, Z. A., Wang, J., Warner, F., Byrne, A. M., and Chilvers, M. I. (2019). Predicting soybean yield and sudden death syndrome development using at-planting risk factors. Phytopathology.

Roy, K. W., Lawrence, G. W., Hodges, H. H., McLean, K. S., and Killebrew, J. F. (1989). Sudden death syndrome of soybean: *Fusarium solani* as incitant and relation of *Heterodera glycines* to disease severity. Phytopathology, 79, 191-197.

Reinprecht, Y., Luk-Labey, S. Y., Larsen, J., Poysa, V. W., Yu, K., Rajcan, I., et al. (2009). Molecular basis of the low linolenic acid trait in soybean EMS mutant line RG10. Plant Breeding, 128, 253-258.

Rennie, B. D., Zilka, J., Cramer, M. M., and Beversdorf, W. D. (1988). Genetic analysis of low linolenic acid levels in the soybean line PI 361088B. Crop Science, 28, 655-657.

Rennie, B. D., and Tanner, J. W. (1989). Fatty acid composition of oil from soybean seeds grown at extreme temperatures. Journal of the American Oil Chemists' Society, 66, 1622-1624.

Ross, A. J., Fehr, W. R., Welke, G. A., and Cianzio, S. R. (2000). Agronomic and seed traits of 1%-linolenate soybean genotypes. Crop Science, 40, 383-386.

Rupe, J. C., Sabbe, W. E., Robbins, R. T., and Gbur, E. E. (1993). Soil and plant factors associated with sudden death syndrome of soybean. Journal of Production Agriculture, 6, 218-221.

Sandhu, D., Alt, J. L., Scherder, C. W., Fehr, W. R., and Bhattacharyya, M. K. (2007). Enhanced oleic acid content in the soybean mutant M23 is associated with the deletion in the *Fad2-1a* gene encoding a fatty acid desaturase. Journal of the American Oil Chemists' Society, 84, 229-235.

Scherm, H., Yang, X. B., and Lundeen, P. (1998). Soil variables associated with sudden death syndrome in soybean fields in Iowa. Plant Disease, 82, 1152-1157.

Schlueter, J. A., Dixon, P., Granger, C., Grant, D., Clark, L., Doyle, J. J., et al. (2004). Mining EST databases to resolve evolutionary events in major crop species. Genome, 47, 868-876.

Schuster, I., Abdelnoor, R. V., Marin, S. R. R., Carvalho, V. P., Kiihl, R. A. S., Silva, J. F. V., et al. (2001). Identification of a new major QTL associated with resistance to soybean cyst nematode (*Heterodera glycines*). Theoretical and Applied Genetics, 102, 91-96.

Sedivy, E. J., Wu, F., and Hanzawa, Y. (2017). Soybean domestication: the origin, genetic architecture and molecular bases. New Phytologist, 214, 539-553.

Shah, S., and Xin, Z. (1997). Overexpression of the FAD3 Desaturase Gene in a Mutant of Arabidopsis. Plant Physiology, 114(4), 1533-1539.

Song, Q., Hyten, D. L., Jia, G., Quigley, C. V., Fickus, E. W., Nelson, R. L., et al. (2013). Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. PloS One, 8, e54985.

Song, Q., Hyten, D. L., Jia, G., Quigley, C. V., Fickus, E. W., Nelson, R. L., et al. (2015). Fingerprinting soybean germplasm and its utility in genomic research. G3: Genes, Genomes, Genetics, 5, 1999-2006.

Srour, A., Afzal, A. J., Blahut-Beatty, L., Hemmati, N., Simmonds, D. H., Li, W., et al. (2012). The receptor like kinase at *Rhg1-a/Rfs2* caused pleiotropic resistance to sudden death syndrome and soybean cyst nematode as a transgene by altering signaling responses. BMC Genomics, 13, 368.

Stijšin, D., Luzzi, B. M., Ablett, G. R., and Tanner, J. W. (1998). Inheritance of low linolenic acid level in the soybean line RG10. Crop Science, 38, 1441-1444.

Swaminathan, S., Abeysekara, N. S., Liu, M., Cianzio, S. R., and Bhattacharyya, M. K. (2016). Quantitative trait loci underlying host responses of soybean to *Fusarium virguliforme* toxins that cause foliar sudden death syndrome. Theoretical and Applied Genetics, 129, 495-506.

Swaminathan, S., Abeysekara, N. S., Knight, J. M., Liu, M., Dong, J., Hudson, M. E., et al. (2018). Mapping of new quantitative trait loci for sudden death syndrome and soybean cyst nematode resistance in two soybean populations. Theoretical and Applied Genetics, 131, 1047-1062.

Sweeney, D. W., Carrero-Colón, M., and Hudson, K. A. (2017). Characterization of new allelic combinations for high-oleic soybeans. Crop Science, 57, 611-616.

Takagi, Y., and Rahman, S. M. (1996). Inheritance of high oleic acid content in the seed oil of soybean mutant M23. Theoretical and Applied Genetics, 92, 179-182.

Tan, R., Serven, B., Collins, P. J., Zhang, Z., Wen, Z., Boyse, J. F., et al. (2018). QTL mapping and epistatic interaction analysis of field resistance to sudden death syndrome (*Fusarium virguliforme*) in soybean. Theoretical and Applied Genetics, 131, 1729-1740.

Tan, R., Collins, P. J., Wang, J., Wen, Z., Boyse, J. F., Laurenz, R. G., et al. (2019). Different loci associated with root and foliar resistance to sudden death syndrome (*Fusarium virguliforme*) in soybean. Theoretical and Applied Genetics, 132, 501-513.

Tewoldemedhin, Y. T., Lamprecht, S. C., Geldenhuys, J. J., and Kloppers, F. J. (2014). First report of soybean sudden death syndrome caused by *Fusarium virguliforme* in South Africa. Plant Disease, 98, 569-569.

Tewoldemedhin, Y. T., Lamprecht, S. C., Vaughan, M. M., Doehring, G., and O'Donnell, K. (2017). Soybean SDS in South Africa is caused by *Fusarium brasiliense* and a novel undescribed Fusarium sp. Plant Disease, 101, 150-157.

Thapa, R., Carrero-Colon, M., Crowe, M., Gaskin, E., and Hudson, K. (2016). Novel *FAD2–1A* alleles confer an elevated oleic acid phenotype in soybean seeds. Crop Science, 56, 226-231.

Valliyodan, B., Qiu, D., Patil, G., Zeng, P., Huang, J., Dai, L., et al. (2016). Landscape of genomic diversity and trait discovery in soybean. Scientific Reports, 6, 23598.

Vierling, R. A., Faghihi, J., Ferris, V. R., and Ferris, J. M. (1996). Association of RFLP markers with loci conferring broad-based resistance to the soybean cyst nematode (*Heterodera glycines*). Theoretical and Applied Genetics, 92, 83-86.

Vuong, T. D., Sleper, D. A., Shannon, J. G., and Nguyen, H. T. (2010). Novel quantitative trait loci for broad-based resistance to soybean cyst nematode (*Heterodera glycines* Ichinohe) in soybean PI 567516C. Theoretical and Applied Genetics, 121, 1253-1266.

Vuong, T. D., Sleper, D. A., Shannon, J. G., Wu, X., and Nguyen, H. T. (2011). Confirmation of quantitative trait loci for resistance to multiple-HG types of soybean cyst nematode (*Heterodera glycines* Ichinohe). Euphytica, 181, 101-113.

Westphal, A., Li, C., Xing, L., McKay, A., and Malvick, D. (2014). Contributions of *Fusarium virguliforme* and *Heterodera glycines* to the disease complex of sudden death syndrome of soybean. PLoS One, 9, e99529.

Wang, D., Diers, B. W., Arelli, P. R., and Shoemaker, R. C. (2001). Loci underlying resistance to race 3 of soybean cyst nematode in *Glycine soja* plant introduction 468916. Theoretical and Applied Genetics, 103, 561-566.

Wang, X., Chen, L., and Ma, J. (2019a). Genomic introgression through interspecific hybridization counteracts genetic bottleneck during soybean domestication. Genome Biology, 20, 22.

Wang, J., Sang, H., Jacobs, J. L., Oudman, K. A., Hanson, L. E., and Chilvers, M. I. (2019b). Soybean sudden death syndrome causal agent *Fusarium brasiliense* present in Michigan. Plant Disease, 103, 1234-1243.

Webb, D. M., Baltazar, B. M., Rao-Arelli, A. P., Schupp, J., Clayton, K., Keim, P., et al. (1995). Genetic mapping of soybean cyst nematode race-3 resistance loci in the soybean PI 437.654. Theoretical and Applied Genetics, 91, 574-581.

Webb, D. M. (2003). U.S. Patent No. 6,538,175. Washington, DC: U.S. Patent and Trademark Office.

Wen, Z., Tan, R., Yuan, J., Bales, C., Du, W., Zhang, S., et al. (2014). Genome-wide association mapping of quantitative resistance to sudden death syndrome in soybean. BMC Genomics, 15, 809.

Wen, L., Yuan, C., Herman, T. K., and Hartman, G. L. (2017). Accessions of perennial Glycine species with resistance to multiple types of soybean cyst nematode (*Heterodera glycines*). Plant Disease, 101, 1201-1206.

Weisemann, J. M., Matthews, B. F., and Devine, T. E. (1992). Molecular markers located proximal to the soybean cyst nematode resistance gene, *Rhg4*. Theoretical and Applied Genetics, 85, 136-138.

White, H. B., Quackenbush, F. W., and Probst, A. H. (1961). Occurrence and inheritance of linolenic and linoleic acids in soybean seeds. Journal of the American Oil Chemists' Society, 38, 113-117.

Winter, S. M., Shelp, B. J., Anderson, T. R., Welacky, T. W., and Rajcan, I. (2007). QTL associated with horizontal resistance to soybean cyst nematode in *Glycine soja* PI464925B. Theoretical and Applied Genetics, 114, 461-472.

Wu, X., Blake, S., Sleper, D. A., Shannon, J. G., Cregan, P., and Nguyen, H. T. (2009). QTL, additive and epistatic effects for SCN resistance in PI 437654. Theoretical and Applied Genetics, 118, 1093-1105.

Xing, L., and Westphal, A. (2013). Synergism in the interaction of *Fusarium virguliforme* with *Heterodera glycines* in sudden death syndrome of soybean. Journal of Plant Diseases and Protection, 120, 209-217.

Yue, P., Arelli, P. R., and Sleper, D. A. (2001a). Molecular characterization of resistance to *Heterodera glycines* in soybean PI 438489B. Theoretical and Applied Genetics, 102, 921-928.

Yue, P., Sleper, D. A., and Arelli, P. R. (2001b). Mapping resistance to multiple races of *Heterodera glycines* in soybean PI 89772. Crop Science, 41, 1589-1595.

Zhang, J., Singh, A., Mueller, D. S., and Singh, A. K. (2015). Genome-wide association and epistasis studies unravel the genetic architecture of sudden death syndrome resistance in soybean. The Plant Journal, 84, 1124-1136.

Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, W., et al. (2015). Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. Nature Biotechnology, 33, 408-414.

CHAPTER 2

GENETIC MAPPING OF RESISTANCE TO SOYBEAN SUDDEN DEATH SYNDROME

I. Abstract

Soybean sudden death syndrome (SDS) is a disease caused by a soil-borne fungus, *Fusarium virguliforme*. In addition to seed treatments, disease resistant varieties are the cornerstone of SDS management. Many genetic studies have attempted to identify genes responsible for the quantitative host resistance to SDS. Three recombinant inbred line (RIL) populations were evaluated for foliar SDS resistance at a naturally infested field site in Decatur, MI during the 2014 and 2015 growing seasons. Lines were evaluated for disease severity (DS) on a 1-9 scale, disease incidence (DI) as an estimate of the percentage of plants with symptoms per plot, and disease index (DX) as a metric which integrates DS and DI. Phenotypic data was spatially adjusted to account for uneven pathogen distribution in the naturally infested field. A subset of RILs from each population were genotyped with the SoySNP6K Illumina Infinium BeadChip. Linkage maps unique to each population were constructed using JoinMap ver. 2. Composite interval mapping was performed using WinQTLCartographer ver. 2.5. Three quantitative trait loci (QTL) were identified across multiple years and/or populations. One QTL on Chromosome Gm10 appeared to be in close proximity to the *E2* maturity locus. Kompetitive Allele Specific Primers (KASP from LGC Genomics) were developed to identify recombination events within this QTL region in the selfed progeny of a residual heterozygous line (RHL). The RHL progeny with recombination events were evaluated in the field site in Decatur, MI to narrow the QTL region responsible for SDS resistance. Another QTL identified was on Chromosome Gm18, in a region which has been demonstrated to provide SCN and SDS resistance in many studies (*rhg1/Rfs2*).

II. Introduction

Soybean (*Glycine max* (L.) Merr.) rivals maize as the most widely grown crop in the U.S., as more acreage of soybean was planted than maize in 2017 and 2018 (National Agricultural Statistics Service 2017, 2018). Soybean grain is composed of approximately 18% oil and 38% protein by weight. The oil is used in processed food, cooking oil, and industrial applications, and the protein meal is a critical component of animal feed, especially for poultry and swine.

Soybean sudden death syndrome (SDS) is an emerging disease threatening soybean production in many continents. The disease was first reported in Arkansas in 1971 and has since spread to nearly all soybean producing regions of North America. The disease is caused by soil-borne fungi from clade 2 of the *Fusarium solani* species complex. In South America, *F. virguliforme*, *F. brasiliense*, *F. crassistipitatum*, and *F. tucumaniae* are causal agents of SDS (Aoki et al., 2005; Aoki et al., 2012). In South Africa, *F. virguliforme* and *F. brasiliense* have been reported to be causal agents (Tewoldemedhin et al., 2014; Tewoldemedhin et al., 2017). In North America *F. virguliforme* was the only reported causal agent of SDS, but new evidence includes *F. brasiliense* as another causal agent of SDS in Michigan (Wang et al., 2019).

Most severe yield losses due to SDS occur in years with cool, wet springs. Yield losses due to SDS range from 21 – 71 million bushels per year, or $185 - $625 million per year in the U.S. and Canada (Allen et al., 2017). In 2010 and 2014, SDS caused the second and third most yield loss of any soybean disease, respectively (Allen et al., 2017). The fungus infects the roots early in the growing season and can cause reduced vigor, and root rot. Later in the season, *F. virguliforme* in the root produces toxins which are translocated to the above ground plant tissues (Brar et al., 2011). During pod-fill (R6 growth stage), foliar interveinal chlorosis and necrosis

can be observed. Sometimes light-blue patches of *F. virguliforme* sporodochia are visible on the taproot near the soil line. In severe conditions, foliar symptoms can be seen in earlier growth stages, which can result in complete defoliation (Hartman et al., 2015).

Common management strategies for SDS include fungicide seed treatments and soybean cyst nematode (SCN) management. Since 2014, fluopyram has been commercially available (BASF, ILeVO®) as a seed treatment to reduce yield losses due to SDS (Wang et al., 2017). Fluopyram seed treatment reduces SDS foliar symptoms by 30-41% and decreases yield loss by 6-11% (Kandel et al., 2016; Kandel et al., 2017).

There is mixed evidence that SDS may be connected to SCN, and that SCN management may help reduce SDS. One study demonstrated that different SCN resistance sources had significant effects on SDS foliar symptoms for three years at six locations in Illinois, Indiana, Iowa, Michigan, and Ontario, Canada (Kandel et al., 2017). The study determined that SDS severity was greatest when there was no SCN resistance. This study also identified a positive correlation between SCN egg counts and SDS foliar symptoms (Kandel et al., 2017). Other evidence of SDS and SCN pleiotropy includes studies by Tan et al., (2018). This study mapped SDS resistance and found an epistatic resistance mechanism in regions overlapping with *rhg1* and *Rhg4*. The pedigree of the resistance source used in this study includes 'Peking', a line which has been demonstrated to derive SCN resistance from an epistatic interaction of *rhg1* and *Rhg4*. Additionally, microscopy analysis has demonstrated that an GFP-tagged *F. virguliforme* infects that surface and interior of *C. elegans* (Mitch Roth, personal communication). If this phenomenon can be confirmed with SCN, a solid mechanism of co-infection can support synergistic and, thus, a pleiotropic model of SDS and SCN resistance.

Recently, a study examined two-, three-, and four-year rotation schedules and their effects on yield loss due to SDS. Extended rotation sequences coupled with the use of organic matter amendments have been demonstrated to be effective at reducing *F. virguliforme* inoculum in the soil (Leandro et al., 2018). The study also demonstrated that a corn/soybean two-year rotation did not significantly reduce *F. virguliforme* inoculum. The authors speculate that oats, which were a part of extended rotations, may be key to reducing *F. virguliforme* inoculum (Leandro et al., 2018). While the extended rotation work is promising, extended rotation schedules have yet to be widely adopted.

Although, fluopyram seed treatment and long-term crop rotations are effective, the development of resistant varieties offers another tool for farmers to use towards reducing yield losses caused by SDS. Compared with seed treatments, resistant varieties are acceptable in both conventional and organic production systems. While long-term crop rotations can be effective, short-term rotations are more widely practiced due to economic factors. Resistant varieties fit within a short-term rotation schedule. In short, planting SDS resistant varieties is critical towards reducing yield losses due to SDS, especially in organic and short-term rotation cropping systems.

In order to develop disease resistant varieties, breeders need to identify genes which are responsible for conferring resistance to the target disease. Once a gene is identified as conferring resistance, or a marker linked to said gene, breeders can select for that gene, as an indirect method of selecting for field resistance. This system is commonly referred to as marker assisted selection (MAS) (Collins et al., 2018). Utilization of MAS greatly improves the efficiency of disease resistance breeding, as genetic tests are often cheaper and more consistent than field resistance screening. However, in order to implement MAS for a specific disease, the gene, or linked markers, must be identified as conferring resistance to that disease.

Many quantitative trait loci (QTL) mapping studies have attempted to identify the loci responsible for SDS resistance. A 2015 review by David Lightfoot summarized breeding work for SDS resistance. The review summarizes eighteen resistance loci (*Rfs1 – Rfs18*) that have been identified and confirmed. These loci include both root and foliar resistance. Root resistance is categorized by root rot and poor root development in the presence of *F. virguliforme*. Foliar resistance is categorized with interveinal necrosis and chlorosis, pod abortion, and early maturity. Foliar resistance involves a mechanism to block toxin translocation from the roots. Several studies indicate that root and foliar resistance are controlled by different genetic mechanisms (Kazi et al., 2008; Tan et al., 2019). Some evidence suggests that foliar resistance is more important at reducing yield losses. The literature also suggests that SDS resistance and SCN resistance may be linked. It is unclear if genetic pleiotropy or independent linked genes are responsible for this observation. The cloning of *GmRLK18-1*, (Wm82.a1.v1 Glyma18g02680) at *Rfs2* (closely linked but independent of r*hg1*) seems to suggest that they are independent, closely linked genes (Srour et al., 2012). However, in the greenhouse assay demonstrating the transgenic SDS resistance, the non-transgenic control without *F. virguliforme* showed moderate foliar disease severity. These sets of data question the validity of the greenhouse assay being used, as an assay without pathogen should not report moderate foliar symptoms. Unfortunately, field trials to confirm transgenic resistance of *Rfs2* failed due to extreme insect herbivory. Additionally, the *GmRLK18-1* was removed from the second assembly of the reference genome.

A 2018 review by Chang et al., proposed a novel nomenclature for SDS-resistance loci which improved upon the 2015 Lightfoot review. The Chang review used a higher reproducibility standard for SDS resistance loci identified in QTL mapping studies, and proposed an "*Rfv*" nomenclature for ten loci associated with SDS resistance. The standard used required

loci to be identified by at least three publications. Chang et al., (2018) also identified literature evidence for disease synergism between SDS and SCN. In short, the authors concluded that additional research is necessary to determine whether pleiotropy has a basis for SDS and SCN resistance.

While phenotyping in environments like growth chambers and greenhouses provide uniformity and the flexibility to phenotype for resistance anytime of the year, the results can often fail to translate to the field. Alternatively, inocula can be planted along with seeds in an artificially infested field, or trials can be conducted in a field with a natural *F. virguliforme* infestation. Like growth chamber and greenhouse studies, artificial inocula adds variables (such as inocula rate, inocula depth, and pathogen isolate) which may confound the applicability of the results to natural field conditions experienced by breeders and farmers. While field phenotyping in a naturally infested field avoids many of these problems, it is complicated by a lack of field uniformity and micro-environmental differences within a field. A potential solution to increase accuracy of field phenotype data is to adjust phenotype data with a spatial data adjustment. Natural field conditions ensure that the results will translate to the farmers' fields, and spatial adjustment reduces the environmental variance which can reduce the statistical power to detect genetic variance.

Spatial data adjustment can account for uneven pathogen distribution within a field, reducing the environmental variance in an experiment. The mvngGrAd R package was developed to minimize environmental variation for plant breeding field trials, and maximize heritability values (Technow, 2011). This package allows the user to spatially adjust their phenotypic data by using the phenotypic data of neighboring plots in a specified grid. The size and shape of the grid can be customized by the user in the 0^{o}, 90^{o}, 180^{o}, and 270^{o} directions, as

well as the other nearby plots. In short, plots within the specified grid are averaged into a moving mean value (x_i) which is used in the following formula to calculate a spatially adjusted phenotypic value for each plot:

$$p_{i,adj} = p_{i,obs} - b(x_i - \ddot{x})$$

Where $p_{i,obs}$ is the observed phenotype, $p_{i,adj}$ is the spatially adjusted phenotype, b is a regression coefficient, x_i is the moving mean described above, and $\ddot{x}$ is the mean of all x_i (Technow, 2011).

However, spatial data adjustment may lead to an overfitting of the data. In order to minimize overfitting and determine the optimum grid for a dataset, several grid types were tested and the adjusted data with the highest heritability should be chosen for the spatial adjustment. This package has been previously used to adjust phenotypic data for stem rust resistance in barley due to uneven infection rates across the test fields (Nice et al., 2017).

The goal of this study was to identify QTL associated with resistance to SDS from the breeding lines E09014 and E09088. These lines have been previously observed to be resistant to SDS based on field data, but the genes underlying their resistance are unknown.

III. Methods

i. Population development

Three recombinant inbred line (RIL) populations which segregated for SDS resistance were developed by crossing an SDS resistant parent with a suspected SDS susceptible parent. All parents were breeding lines from the Michigan State University Soybean Breeding Program. Population ‘PA’ was composed of 269 RILs and was developed from the cross E09088 (resistant) x E12901 (moderately resistant); Population ‘PB’ was composed of 124 RILs and was developed from the cross E05226-T (susceptible) x E09014 (resistant); Population ‘PC’ was

composed of 226 RILs and was developed from the cross E05226-T (susceptible) x E09088 (resistant).

While the intention was for all populations to be derived from one resistant and one susceptible parent, E12901, which was suspected to be susceptible to SDS, showed moderate resistance upon further field evaluation. As a consequence of this misidentification, the PA population had lower foliar symptom scores and a lower range of foliar symptoms scores compared to populations PB and PC (Table 2.1). It is also of note that the resistant parents from all three populations, E09014 and E09088, both contain SCN resistance derived from PI 88788.

Table 2.1. Unadjusted SDS resistance data for Populations PA, PB, and PC*. Data is shown for RILs and parents collected in Decatur, MI from 2014 and 2015. The percentage of plots with no symptoms is presented as escape percentage.*

		2014					2015				
Pop	**Trait**	**RIL mean**	**RIL range**	**Escapes[a]**	**Res. parent[b]**	**Sus. parent[c]**	**RIL mean**	**RIL range**	**Escapes[a]**	**Res. parent[b]**	**Sus. parent[c]**
PA	DS	1.03	0 - 4	54.65%	0.5	0.5	0.38	0 - 3	86.32%	0	0.625
	DI	13.72	0 - 71.25		1.67	1.67	2.48	0 - 40		0	5
	DX	4.04	0 - 31.8		0.28	0.28	0.85	0 - 18.75		0	1.39
PB	DS	1.42	0 - 4	50.10%	0	3.5	1.56	0 - 4.625	48.95%	0.25	4.25
	DI	18.67	0 - 75		0	36.67	19.61	0 - 86.25		1.25	57.5
	DX	6.64	0 - 33.96		0	15.28	7.74	0 - 43.19		0.56	21.25
PC	DS	1.82	0 - 4.375	30.20%	0	2.25	2.13	0 - 4.75	35.11%	0.375	4.5
	DI	23.97	0 - 80		0	30	28.99	0 - 83.75		1.25	66.25
	DX	8.22	0 - 34.24		0	12.08	11.78	0 - 40.76		0.21	35.42
			Total	44.78%				**Total**	57.84%		

[a] Escapes indicate phenotypic values of zero, or no SDS symptoms

[b] Resistant Parent for Population PA and PC is E09088, for Population PB is E09014

[c] Susceptible Parent for Population PA is E12901, for Population PB and PC is E05226-T

All populations were advanced via single seed descent until the F4 generation. Single F4 plants were grown in the field and threshed individually to generate the F4-derived RILs. Prior to initial mapping, these lines were planted, and bulk harvested to advance from F4:5 to F4:6. F4:6

lines were used for the first year of field evaluations. Lines were bulk harvested from F4:6 to advance to F4:7. F4:7 lines were used for the second year of field evaluations.

ii. Phenotyping

RILs for all three populations, and parents, were planted in four replications of three-meter, single row plots (1 meter spacing between plots) at a field site near Decatur, MI for two years. This site has a natural infestation of *F. virguliforme* and high levels of soybean cyst nematodes. In 2014, the RILs were F4:6, and in 2015 the RILs were F4:7. These plots were maintained using conventional soybean management to control weeds. Overhead irrigation was applied to plots to ensure environmental conditions favorable for SDS symptom development.

Visual rating of research plots was conducted during pod-fill (R6 growth stage). Disease severity (DS) was determined based on rating scale developed by Chilvers and Sevren. This is a visual rating scale based on the approximate percentage of leaf area showing necrotic and/or chlorotic symptoms on symptomatic plants within the plot. Disease Incidence (DI) is a measure of the percentage of plants displaying symptoms within a given plot. DI ranges from 0 to 100 in increments of 5. Disease Index (DX) combines the above traits into an overall SDS resistance metric (DX = (DS/9) * DI), which takes into account the severity and prevalence of SDS symptoms.

iii. Spatial data adjustment

The R package MvngGrAd was used to adjust phenotypic values based on neighboring plots. The purpose of this adjustment was to account for uneven pathogen pressure in the field, which

manifested in sections of the field where plots showed little to no symptoms. This resulted in all raw phenotypic distributions being left-skewed. The R package allows the user to specify the number and spatial pattern of neighboring plots used to adjust phenotypic values (Technow, 2011). Four different patterns were selected to adjust phenotypic values. Broad sense heritability values for all phenotypic traits were calculated before and after spatial data adjustment. The formulas for broad sense heritability were from Nyquist and Baker (1991). PROC GLM procedure in SAS 9.4 (SAS Institute Inc., Cary, NC, USA was used to apply these formulae to the data. An increase in heritability after spatial data adjustment indicated that the spatial data adjustment was successful in reducing the environmental variation caused by uneven pathogen pressure. The change in heritability also allowed comparison of multiple patterns, and selection of the best pattern to maximize the proportion of total variance explained by genetic variance, thus increasing the power for QTL detection.

Four spatial data adjustments patterns were selected to adjust phenotypic values based on values of neighboring plots. Adjustment 'alpha' was selected to account for neighboring plots within 2 ranges (6m), 2 passes (1.5m), and immediately diagonally adjacent. Adjustment 'beta' accounted for neighboring plots within 1 range (3m), 4 passes (3m), and immediately diagonally adjacent. Adjustment 'gamma' accounted neighboring plots within 2 ranges (6m), 4 passes (3m), and immediately diagonally adjacent plots. Adjustment 'delta' accounted for neighboring plots within 1 range (3m) and 4 passes (3m) (Figure 2.1). The pattern which contributed the largest increase in heritability, averaged across years and populations, was selected for spatial data adjustment.

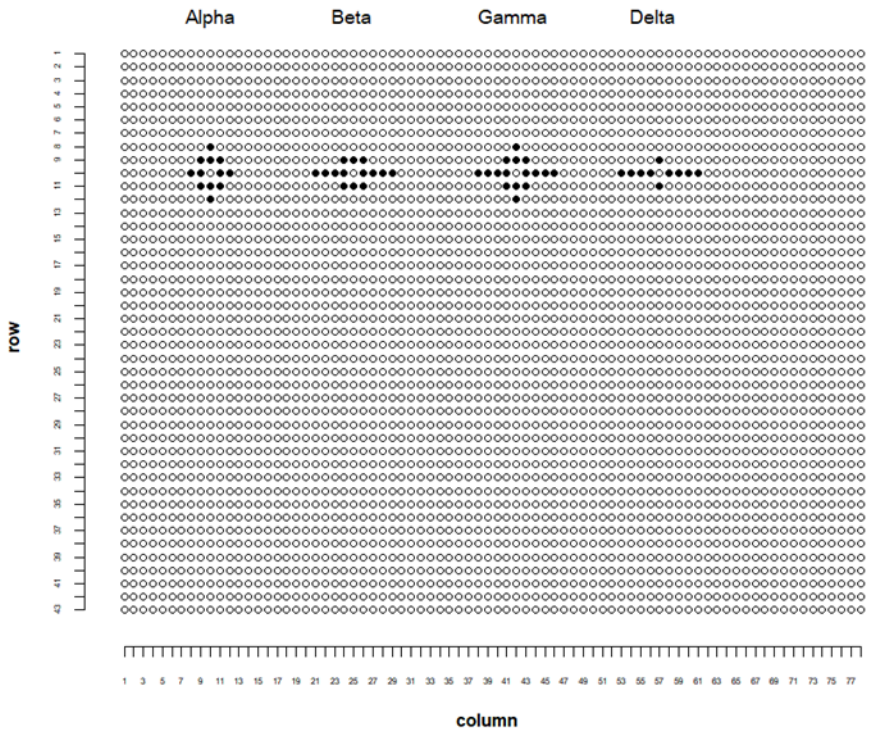


Figure 2.1. *Spatial adjustment grid patterns.* This figure depicts the four grid patterns used for spatial adjustment of phenotypic data. These grids were applied using the R package MvngGrAd (Technow, 2011) to calculate a moving mean value used to calculate a spatially adjusted phenotypic value.

iv. Genotyping and linkage map construction

For the three RIL populations and the parents, young leaf tissue from ten plants from each F4:6 line was bulked and DNA was extracted using a modified CTAB extraction method. DNA concentration was diluted to 50 nanograms/microliter. The parents, and a subset of each population (116 - 125 lines per population) were genotyped with the SoySNP6K Illumina Infinium BeadChip. SNP calls from the SoySNP6k Illumina Infinium BeadChip was interpreted in Genome Studio.

Linkage maps were constructed separately for each RIL population. SNP data was filtered to remove monomorphic markers based on the parental genotypes. SNP markers were filtered out if parental genotypes were missing, heterozygous, or shared between the parents. Each linkage map was constructed with 1321-2228 markers after filtering. The data was used to construct three different linkage maps for each population using JoinMap (ver. 2), using a LOD of 3 to group markers and the regression algorithm to determine marker orders. Linkage groups with less than five markers were not included in the linkage maps. The linkage map positions largely corresponded with the physical positions of makers, with some chromosomes were represented by two distinct linkage groups corresponding to different chromosome arms.

v. Quantitative trait loci (QTL) detection

The linkage maps were imported into Windows QTL Cartographer (ver. 2.5) along with the mean of four replications of spatially adjusted phenotype data (DS, DI, and DX) for the years 2014 and 2015. Composite interval mapping (CIM) was conducted using a walk speed of 0.5cM, forward and backward selection of markers as cofactors, and the Zmapqtl 6 model. The LOD threshold for QTL significance was determined for each population using a 1000 permutation test with alpha = 0.05. Flanking markers for each QTL were identified using a +/- 1LOD range from the maximum LOD of the QTL peak.

vi. Residual heterozygous line analysis

To confirm the effect of QTL_10, the PB RIL population was screened with seven KASP custom SNP markers designed to cover a 5.5 Mb window encompassing the QTL (Table 2.5). The intention was to identify a residual heterozygous line (RHL) from the initial mapping population

to develop fine mapping lines. PB-052 was selected as the RHL, as it was heterozygous across the 5.5 Mb window, 84.6% homozygous across the total genome, and homozygous for the susceptible allele at QTL_18. It was important to select a line with the susceptible allele at QTL_18 because the effect of resistance from QTL_18 could mask the resistance from QTL_10. Seeds from PB-052 were planted in the greenhouse, tissue was collected from young leaves, DNA was extracted using the same CTAB protocol, and the PB-052-derived lines were screened with the same set of KASP markers. Based on the KASP marker data, five F7 RHL-derived mapping lines were selected with recombination breakpoints within the 5.5 Mb window. Seed was increased for these lines in 2016, lines were planted in the Decatur field for the 2017 growing season with four replicates for each line, and lines were rated for foliar symptoms.

IV. Results

i. Phenotyping

In 2014 and 2015, each plot was rated for resistance to SDS based on disease severity (DS), disease incidence (DI), and disease index (DX). In 2014, mean DS for RILs were 1.03, 1.42, and 1.82 for populations PA, PB, and PC, respectively. Mean DI values were 13.72, 18.67, and 23.97 for populations PA, PB, and PC. Mean DX values were 4.04, 6.64, and 8.22 for populations PA, PB, and PC, respectively. In 2015, mean DS values were 0.38, 1.56, and 2.13 for populations PA, PB, and PC, respectively. Mean DI values were 2.48, 19.61, and 28.99 for populations PA, PB, and PC, respectively. Mean DX values were 0.85, 7.74, and 11.78 for populations PA, PB, and PC, respectively. For population PA, phenotypic data between parents was not significantly different. Prior to population development, E12901 was falsely identified as susceptible to SDS. However, data collected in 2014 and 2015 identified E12901 as moderately resistant to SDS.

Despite E09088 being resistant and E12091 being moderately resistant, population PA did segregate for SDS resistance. This confirms the findings of previous studies that SDS is a quantitative trait which is controlled by many loci. The range and mean values for population PA was lower than the other populations, due to resistance being contributed from both parents. These differences were especially pronounced in the 2015 data.

The raw data for each trait and each population x year combination was left-skewed, as field escapes were prevalent in both years (Figure 2.3). Plots with no SDS symptoms accounted for 44.78% of the total number of plots in 2014 and 57.84% of plots in 2015. Within population PA in 2015, disease pressure was so low that 86.32% of plots showed no symptoms. The high rate of escapes is accounted for by an uneven pathogen distribution and/or microclimatic conditions which allows plots within that field to display zero symptoms, regardless of the resistant/susceptible status of the genotype (Figure 2.2). Skewed phenotypic distributions are not ideal for QTL mapping analysis, so data was transformed based on a spatial analysis of phenotypic data of neighboring plots, in order to reduce the effects of escapes, and increase the proportion of genetic variance captured by the phenotypic data.

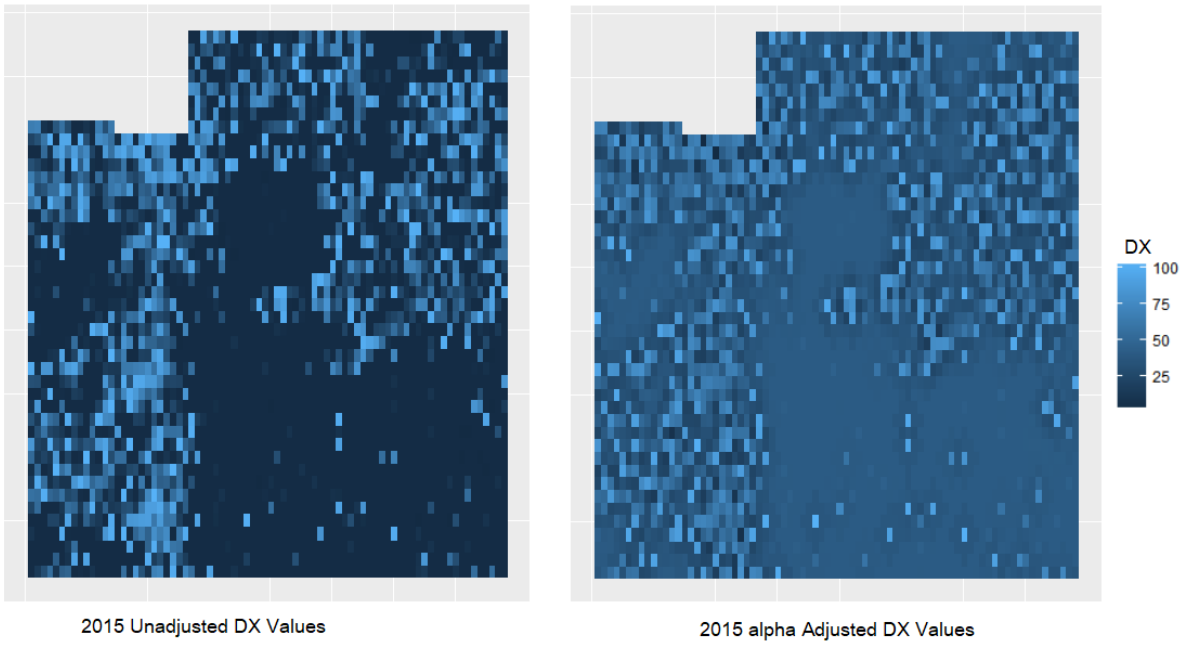


Figure 2.2. *2015 disease index (DX) values before and after alpha adjustment.* This figure depicts a heatmap of 2015 DX values, before and after spatial adjustment. Note the patchiness of DX values, dark values are almost uniform in large sections of the unadjusted DX heat map.

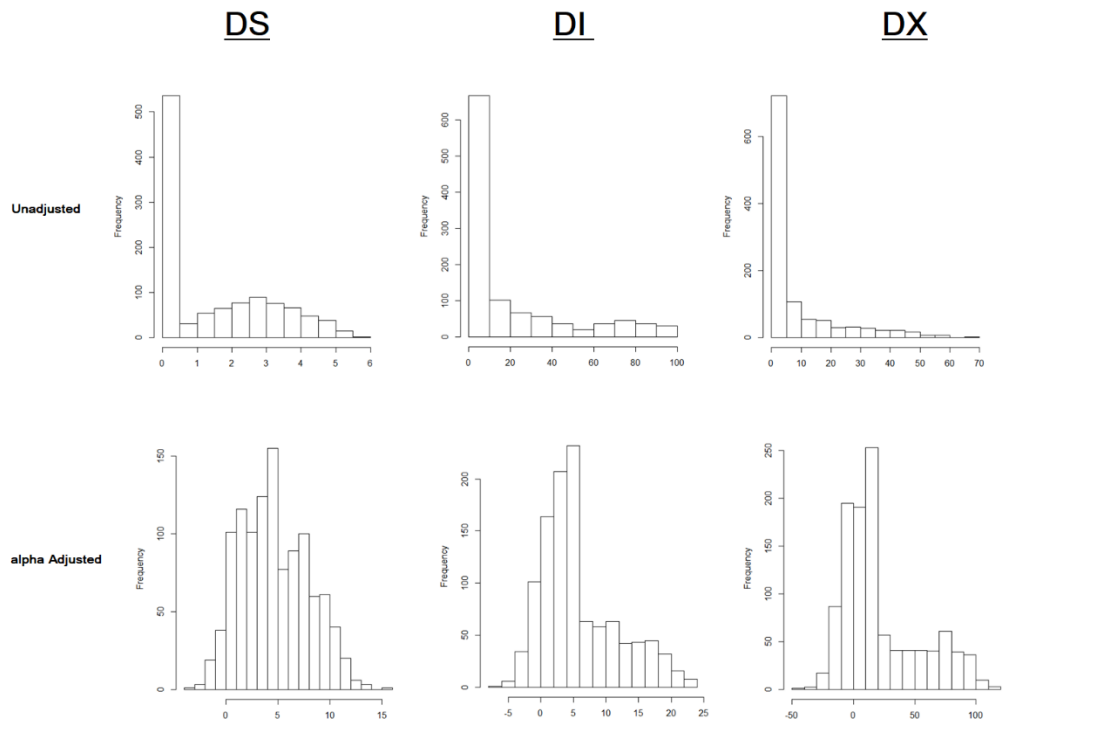


Figure 2.3. *PB 2015 phenotypic distributions before and after alpha adjustment*. This figure illustrates the benefit of spatial adjustment in normalizing phenotypic data. Note the large number of zero values for all traits for all the unadjusted phenotypes.

ii. Heritability

Prior to spatial data adjustment, heritability scores for DS, DI, and DX ranged from 0.43 to 0.56, 0.21 to 0.46, and 0.36 to 0.55, respectively. Overall, population PA had lower heritability scores than the other populations, likely due to less phenotypic variation between the parents which increased the proportion of variance due to the environment relative to the genetic variance. In 88% of cases, the application of spatial data adjustment patterns increased heritability (Table 2.2). Adjustment 'alpha' showed the largest increase in heritability scores in eight of the nine population-trait combinations. The largest increase in heritability was for DI in population PB using pattern alpha, which increased heritability scores from 0.46 to 0.53. Based on the observed

increases in heritability scores, adjustment 'alpha' was selected, as it had the largest average increase in heritability scores, increasing heritability by an average of 15.0% across the three traits (Table 2.2). The alpha adjustment used phenotype scores of the plots 2 ranges to the left and the right, 2 passes towards the front and the back, and the 4 diagonally adjacent plots (Figure 2.1). After the spatial adjustments, adjusted DS, DI, and DX scores fit roughly normal distributions for all population by year combination (Figure 2.3).

***Table 2.2a. Broad sense heritability of disease severity**. Heritability was calculated or disease severity (DS) for Populations PA, PB, and PC for data collected from Decatur, MI in 2014 and 2015*

	Population	PA	PB	PC	Average % Increase
	Original DS	0.39	0.53	0.48	-
Adjustment	**Alpha**	0.48	0.62	0.5	14.28
	Beta	0.48	0.61	0.47	11.43
	Gamma	0.47	0.61	0.48	11.43
	Delta	0.46	0.61	0.47	10.00

***Table 2.2b. Broad sense heritability of disease incidence**. Heritability was calculated for disease incidence (DI) for Populations PA, PB, and PC for data collected from Decatur, MI in 2014 and 2015*

	Population	PA	PB	PC	Average % Increase
	Original DI	0.19	0.45	0.45	-
Adjustment	**Alpha**	0.23	0.53	0.5	15.60
	Beta	0.27	0.51	0.47	14.68
	Gamma	0.25	0.51	0.48	13.76
	Delta	0.26	0.51	0.47	13.76

Table 2.2c. Broad sense heritability of disease index*. Heritability was calculated for disease index (DX) for Populations PA, PB, and PC for data collected from Decatur, MI in 2014 and 2015*

	Population	PA	PB	PC	Average % Increase
	Original DX	0.32	0.51	0.36	-
Adjustment	**Alpha**	0.37	0.61	0.39	15.13
	Beta	0.37	0.6	0.37	12.61
	Gamma	0.36	0.61	0.38	13.45
	Delta	0.35	0.61	0.36	10.92

iii. Linkage maps

Linkage maps were constructed with SNP chip data from 116, 118, and 125 individuals from populations PA, PB, and PC, respectively (Table 2.3). After filtering using parent data and segregation distortion, 2228, 1361, and 1365 markers were polymorphic for populations PA, PB, and PC, respectively. JoinMap was used to group markers into 26, 21, and 20 linkage groups for populations PA, PB, and PC, respectively. A regression algorithm was used to order markers within each linkage group. Linkage groups were composed of 9-156, 8-113, and 21-105 markers per linkage group for PA, PB, and PC respectively. The size of linkage groups ranged from 9.6-209.5, 15.6-250.1, and 62.4-326.3 cM for populations PA, PB, and PC, respectively. Average distance between markers was 1.18, 2.75, and 3.47 cM for populations PA, PB, and PC, respectively.

Table 2.3a. Linkage map for population PA. *The linkage map for population PA was constructed using JoinMap ver. 2.*

Population PA			**Linkage Groups**		**Chromosomes**	
Chromosome	Linkage Group	#Markers	Linkage Group Size (cM)	Average Distance (cM) between Markers	Chromosome Size (cM)	Average Distance (cM) between Markers
Gm01	LG1	109	191.09	1.75	191.09	1.75
Gm02	LG2	118	209.52	1.78	209.52	1.78
Gm03	LG3	107	144.89	1.35	144.89	1.35
Gm04	LG4	64	119.56	1.87	119.56	1.87
Gm05	LG5.1	41	67.57	1.65	140.44	2.13
	LG5.2	25	72.87	2.91		
Gm06	LG6.1	59	125.83	2.13	160.77	1.94
	LG6.2	24	34.94	1.46		
Gm07	LG7	151	131.31	0.87	131.31	0.87
Gm08	LG8.1	136	157.02	1.15	260.32	0.65
	LG8.2	129	31.40	0.24		
	LG8.3	136	71.90	0.53		
Gm09	LG9	98	151.19	1.54	151.19	1.54
Gm10	LG10	65	56.51	0.87	56.51	0.87
Gm11	LG11	117	142.64	1.22	142.64	1.22
Gm12	LG12	62	98.48	1.59	98.48	1.59
Gm13	LG13.1	94	91.03	0.97	100.66	0.98
	LG13.2	9	9.64	1.07		
Gm14	LG14	82	123.73	1.51	123.73	1.51
Gm15	LG15	64	124.64	1.95	124.64	1.95
Gm16	LG16	156	90.87	0.58	90.87	0.58
Gm17	LG17.1	55	61.40	1.12	114.87	1.32
	LG17.2	32	53.47	1.67		
Gm18	LG18	106	163.85	1.55	163.85	1.55
Gm19	LG19	91	103.68	1.14	103.68	1.14
Gm20	LG20	98	145.27	1.48	145.27	1.48
Total	**26**	**2228**	**-**	**-**	**2629.02**	**1.18**

Table 2.3b. Linkage map for population PB. *The linkage map for population PB was constructed using JoinMap ver. 2.*

Population PB			**Linkage Groups**		**Chromosomes**	
Chromosome	Linkage Group	#Markers	Linkage Group Size (cM)	Average Distance (cM) between Markers	Chromosome Size (cM)	Average Distance (cM) between Markers
Gm01	LG01	50	169.22	3.38	169.22	3.38
Gm02	LG02	68	241.38	3.55	241.38	3.55
Gm03	LG03	51	250.15	4.90	250.15	4.90
Gm04	LG04	81	183.97	2.27	183.97	2.27
Gm05	LG05	82	219.12	2.67	219.12	2.67
Gm06	LG06	95	203.71	2.14	203.71	2.14
Gm07	LG07.1	56	92.25	1.65	107.83	1.68
	LG07.2	8	15.58	1.95		
Gm08	LG08	70	236.93	3.38	236.93	3.38
Gm09	LG09	73	191.49	2.62	191.49	2.62
Gm10	LG10	65	221.71	3.41	221.71	3.41
Gm11	LG11	30	187.55	6.25	187.55	6.25
Gm12	LG12	52	144.99	2.79	144.99	2.79
Gm13	LG13	73	111.65	1.53	111.65	1.53
Gm14	LG14	33	127.50	3.86	127.50	3.86
Gm15	LG15	113	151.60	1.34	151.60	1.34
Gm16	LG16	93	241.59	2.60	241.59	2.60
Gm17	LG17	64	227.99	3.56	227.99	3.56
Gm18	LG18	100	204.16	2.04	204.16	2.04
Gm19	LG19	68	192.77	2.83	192.77	2.83
Gm20	LG20	36	133.29	3.70	133.29	3.70
Total	**22**	**1361**	-	-	**3748.59**	**2.75**

Table 2.3c. Linkage map for population PC. *The linkage map for population PC was constructed using JoinMap ver. 2.*

Population PC			**Linkage Groups**		**Chromosomes**	
Chromosome	Linkage Group	#Markers	Linkage Group Size (cM)	Average Distance (cM) between Markers	Chromosome	Linkage Group
Gm01	LG01	76	-	-	228.95	3.01
Gm02	LG02	81	-	-	272.26	3.36
Gm03	LG03	37	-	-	185.81	5.02
Gm04	LG04	84	-	-	219.85	2.62
Gm05	LG05	71	-	-	284.34	4.00
Gm06	LG06	74	-	-	255.47	3.45
Gm07	LG07	63	-	-	164.47	2.61
Gm08	LG08	54	-	-	265.31	4.91
Gm09	LG09	86	-	-	146.38	1.70
Gm10	LG10	92	-	-	322.63	3.51
Gm11	LG11	76	-	-	259.67	3.42
Gm12	LG12	69	-	-	326.05	4.73
Gm13	LG13	76	-	-	211.05	2.78
Gm14	LG14	46	-	-	174.89	3.80
Gm15	LG15	105	-	-	257.49	2.45
Gm16	LG16	51	-	-	203.48	3.99
Gm17	LG17	42	-	-	326.33	7.77
Gm18	LG18	90	-	-	283.74	3.15
Gm19	LG19	71	-	-	283.47	3.99
Gm20	LG20	21	-	-	62.45	2.97
Total	**20**	**1365**	**-**	**-**	**4734.09**	**3.47**

iv. QTL results

Results from the QTL detection are summarized in Table 2.4 and Figure 2.4. In total, 20 QTL were detected at six loci on five chromosomes. Three of the six loci were only detected in a single year and a single population. Three of the loci were detected across multiple populations and multiple years.

Table 2.4. Summary of quantitative trait loci (QTL)*. QTL were detected using linkage maps and SDS resistance data collected from Decatur, MI in 2014 and 2015.*

QTL name	Population x Year x Trait	Chromosome	Peak Physical Position[a] (Mb)	Flanking Markers	Peak LOD	R^2	Resistant Allele Derived from
SDS_04	PA_2014_aDS PA_2014_aDX	Gm04	51,736,487	Gm04_47313368_T_G Gm04_48605014_T_G	5.87	16.00%	E12901
SDS_05	PB_2014_aDX	Gm05	32,832,462	Gm05_31827868_T_G Gm05_33075880_T_C	3.55	5.63%	E05226T
SDS_10	PB_2015_aDI PC_2015_aDS	Gm10	45,152,633	Gm10_44445941_A_G Gm10_44972284_T_C	6.05	10.53%	E05226T
SDS_17	PA_2015_aDI PB_2014_aDI	Gm17	13,875,039	Gm17_13253012_A_G Gm17_37391436_C_T	5.03	14.04%	E09088 (PA) and E09014 (PB)
SDS_18a	PA_2014_aDX PB_2014_all PB_2015_all PC_2014_all PC_2015_aDS PC_2015_aDX	Gm18	1,713,268	Gm18_1112389_C_A Gm18_1861988_T_C	15.66	36.38%	E09088 (PA and PC) and E09014 (PB)
SDS_18b	PC_2014_aDI	Gm18	53,247,366	Gm18_57446256_C_T Gm18_58663461_T_C	3.42	6.21%	E09088

[a] Physical positions are based on Wm82.a2.v1

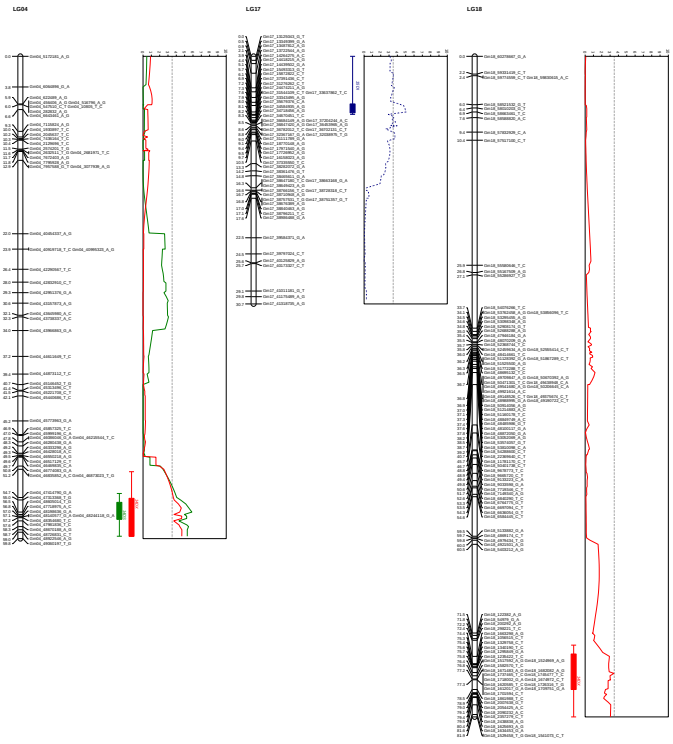


Figure 2.4a. *QTL positions on linkage groups for population PA.* QTL were detected using Windows QTL Cartographer with the composite interval mapping method.

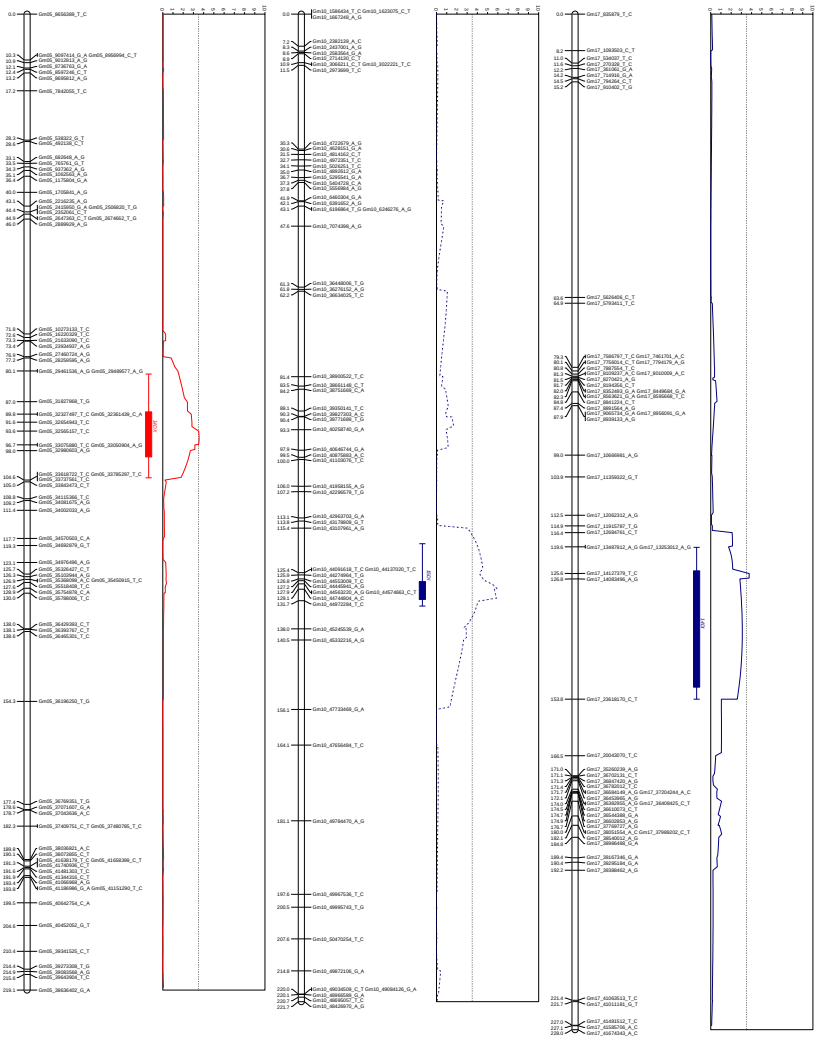
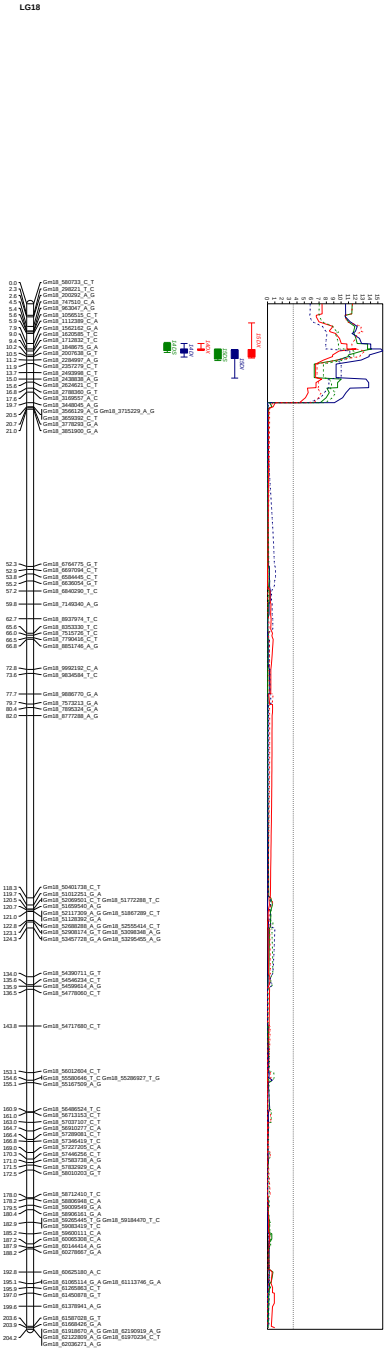


Figure 2.4b. *QTL positions on linkage groups for population PB.* QTL were detected using Windows QTL Cartographer with the composite interval mapping method.

Figure 2.4b. (cont'd)



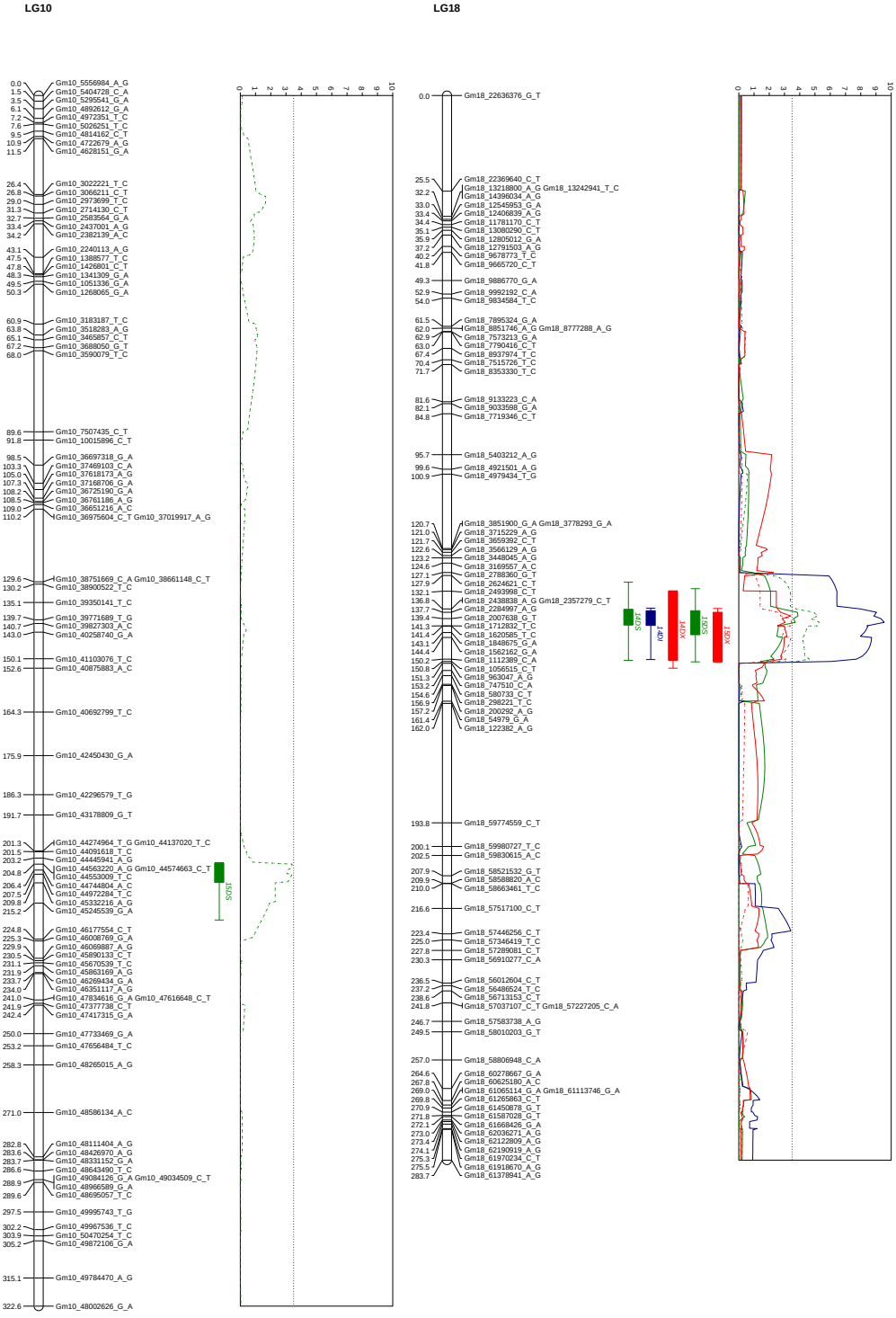


Figure 2.4c. *QTL positions on linkage groups for population PC.* QTL were detected using Windows QTL Cartographer with the composite interval mapping method.

The SDS_04 QTL was detected in 2015 in population PA for both DS and DX. While it does not overlap with any of the *Rfv* loci, it does overlap tightly with SDS9-3 (Njiti and Lightfoot, 2006) and a minor QTL from another SDS mapping study (Tan et al., 2018). The SDS_04 QTL spanned 1.29 Mb based on physical position of markers (50.4 – 51.7 Mb), and accounted for 11.4-16.0% of the phenotypic variance observed. The resistance allele for SDS_04 was contributed by the moderately resistant parent, E12901. The system proposed by Chang et al., 2018 for *Rfv* nomenclature recommends three publications before an *Rfv* name is given. This is the third publication for this QTL.

The SDS_05 QTL was detected in 2014 using DX values for population PB. The SDS_05 QTL spanned 1.25 Mb based on the physical position of the markers (32.1 - 33.3 Mb) and accounted for 5.6% of the phenotypic variance observed. The resistance allele for SDS_05 was contributed by E05226-T, the susceptible parent. SDS_05 is in close proximity to a QTL described by Swaminathan et al 2016 (SDS15-8).

The SDS_10 QTL spanned 0.527 Mb based on the physical position (45.0 - 45.6 Mb) and accounted for 8.7% - 10.5% of the phenotypic variance observed. It was detected in 2015 for DS and DI for populations PC and PB, respectively. In both populations, the resistant allele was derived from E05226-T, the susceptible parent. This region has been previously identified as involved in SDS resistance (Anderson et al., 2015). Interestingly, this region overlaps with the *E2* maturity locus. It may be possible that maturity plays a role in resistance, or that phenotyping is complicated by maturity differences. A correlation test showed significant ($p < 0.05$), but weak ($R = 0.14$) correlation between maturity and the spatially adjusted DX value for PB and PC in 2015. The custom KASP markers developed for this region identified PB-052 as an F7 residual heterozygous line. Field evaluation of five lines derived from PB-052 was conducted in 2017 and

foliar disease scores were calculated at R6 growth stage. The foliar data was compared to the KASP genotypic data and parental KASP data to narrow the region (Table 2.5). In conjunction with the QTL mapping results, the analysis indicated that the causal locus was between marker Gm10_44445941_A_G and marker Gm10_44744804_A_C, a 0.299 Mb region. The E2 locus (Glyma.10g221500) is within this region, in addition to 35 other gene models including a leucine-rich repeat receptor-like kinase (Glyma.10g222600).

Two QTL were detected on Chromosome 17, in close proximity to each other based on physical position. They will both be referred to as SDS_17. SDS_17 was detected in 2015 using DI for population PA. For this population and trait, SDS_17 spanned 24.1 Mb based on the physical positions of the markers (13.0 - 37.1 Mb) and accounted for 14.0% of the phenotypic variance observed. This very large region identified corresponds to a region of the PA linkage map which had a short genetic distance, despite having large physical distance between markers. This may be due to limited recombination in this region, which includes the Chromosome Gm17 centromere (~26 Mb). The resistant allele for SDS_17 in population PA was derived from the resistant parent, E09088. SDS_17 was also detected in 2014 using DI from population PB. It spanned 0.876 Mb and accounted for 4.9% of the phenotypic variance observed. For population PB, the resistant allele came from the resistant parent, E09014. SDS_17 is in close proximity to *Rfv17-01*, which has been identified by many QTL mapping studies to be involved in root resistance to SDS (Kazi et al., 2008; Bao et al., 2015; Lightfoot, 2015). While we measured a foliar resistance trait, it may be that root resistance in populations PA and PB contributed to observable foliar differences.

The SDS_18a QTL spanned 0.750 Mb according to physical position of markers and accounted for up to 36.4% of the phenotypic variance observed. It was detected for all traits (DS,

DI, and DX), in both years, and in all three populations. In all populations, the resistant allele was contributed by the resistant parent, E09088 for PA and PC, and E09014 for PB. This region overlaps with the *rhg1/Rfs2* (*Rfv18-01*) locus, identified by many studies to be important in SDS and SCN resistance. Due to the high r^2 value, this QTL may serve as a useful candidate to develop markers for SDS resistance breeding. Further fine mapping would be necessary to develop markers, but as this QTL overlaps with the r*hg1/Rfs2* locus, we hypothesize that this known locus is responsible for the SDS resistance QTL detected on Gm18. SDS_18b was also detected on Chromosome Gm18, however it was on the opposite arm from *rhg1/Rfs2*. SDS_18b was detected in 2014 using DI for population PC. This QTL spanned 1.21 Mb according to marker physical position and accounted for 6.2% of the phenotypic variance observed. The resistance allele for SDS_18b was contributed by E09088, the resistant parent. SDS_18b is in close proximity to marker ss715631747_C_T (Zhang et al., 2015), marker BARC-024251-04812 (Bao et al., 2015), and QTL SDS4-2 (Njiti et al., 1998), which have been identified to be associated with SDS resistance.

v. RHL results

949 selfed progeny from the RHL 'PB-052' were screened with seven KASP markers spanning the QTL_10 region, and only five high-confidence, homozygous recombinants were identified. Seed from these recombinants was planted at the Decatur, MI field site in 2017 in three replicated plots. DS, DI and DX were measured in the same manner as previously described. Average DX scores and the KASP marker data was used to narrow the QTL_10 region from a 527 kb to a 299 kb region of the soybean genome. Two of the RHL lines were classified as susceptible (DX > 15) and three were classified as resistant (DX < 15). The genotype data

showed that both susceptible lines shared the E05226-T allele (resistant allele donor) at Gm10_44744804_A_C, and all three resistant lines shared the E09014 allele (susceptibility allele) at Gm10_44744804_A_C. This suggests that a recombination breakpoint has occurred between Gm10_44744804_A_C and the causal locus. Specifically, the KASP data moved the QTL right flank from the 45.6 Mb position to the 45.0 Mb position on Gm10. Inspection of the 299 kb region between Gm10_44445941_A_G and Gm10_44744804_A_C on the soybean reference genome revealed 36 gene models, including the gene model corresponding to the *E2* maturity locus (Glyma.10g221500) and a gene model corresponding to a leucine-rich repeat receptor-like protein kinase (Glyma.10g222600), a class of genes known to be associated with disease resistance.

Table 2.5. PB-052 fine mapping. *Fine mapping was conducted to narrow the region detected by QTL analysis. Markers within this region were coded as corresponding to parent E09014 or E05226-T. SDS resistance data was collected from Decatur, MI in 2017 and used to narrow the region.*

Line	17DX	R or S	Maturity Date	Gm10 39827303 A_C	Gm10 40258740 G_A	Gm10 41103076 T_C	Gm10 44274964 T_G	Gm10 44445941 A_G	Gm10 44744804 A_C	Gm10 45332216 A_G
PB-052-017	19.86	S	15-Sep	E09014	E09014	E09014	Unknown	E09014	E05226-T	E05226-T
PB-052-019	17.77	S	9-Sep	E09014	E09014	E09014	Unknown	E09014	E05226-T	E05226-T
PB-052-918A	12.36	R	26-Sep	E09014	E09014	E05226-T	E05226-T	E05226-T	E09014	E09014
PB-052-815A	9.17	R	5-Oct	E05226-T	E05226-T	E05226-T	E05226-T	E05226-T	E09014	E09014
PB-052-874A	5.07	R	26-Sep	E09014	E09014	E05226-T	E05226-T	E05226-T	E09014	E09014

V. Discussion

In summary, six QTL were identified from three RIL populations based on spatially adjusted foliar ratings conducted in a naturally infested field site over a two-year period. QTL_04, QTL_05, and QTL_18b were only detected in a single population during a single year of the study. These are considered low-confidence QTL, as they were not detected in multiple years and/or populations. However, these QTL all overlap with other SDS resistance QTL reported in the literature. QTL_10, QTL_17, and QTL_18a are considered high confidence QTL, as they are supported by multiple years and/or populations. These QTL have also been previously reported in the SDS resistance QTL literature.

QTL_10 was detected in populations PB and PC during the 2015 season and appears to be localized near the *E2* maturity locus. While fine mapping using RHL-derived lines narrowed the QTL to a 299 kb region, this region contained 36 gene models including *E2*. It is possible that the *E2* locus was detected due to maturity differences confounding phenotyping. A significant, but weak correlation between maturity and DX may indicate that maturity is playing a role in avoidance, or that maturity is confounding accurate resistance ratings. *E2* may be pleiotropic in affecting maturity and resistance, via an avoidance mechanism, based on the correlation, lines with earlier maturity tended to show more resistance. The differences in developmental stage during infection or toxin translocation could be responsible for this pleiotropy, or one of the other 35 genes within the 299 kb region could be responsible for the resistance observed however the resolution of our fine mapping was not able to determine the causal gene. Further research is needed to determine the relationship between SDS resistance and the *E2* maturity locus.

QTL_17 was detected in PA during 2015, and PB during 2014. While the regions identified do not technically overlap, they are nearly adjacent. The lack of overlap may be attributed to the linkage map developed for population PA. Within the QTL_17 region identified from PA a lack of recombination may have caused the short genetic distance to correspond to a large physical distance. The centromere for Chromosome Gm17 is within this region, so the lack of recombination is not surprising, given that recombination rates in peri-centromeric regions have been demonstrated to be low compared to the arms of the chromosomes. The QTL_17 region is supported by other QTL mapping studies for SDS root resistance (Kazi et al., 2008; Bao et al., 2015; Lightfoot et al., 2015). Studies have identified root resistance and foliar resistance as separate genetic mechanisms (Kazi et al., 2008; Tan et al., 2019). However, it is of note that the foliar-based QTL_17 identified in this study seems to be supported in the literature by studies focusing on SDS root resistance. Quantitative PCR (qPCR) tools are now available to accurately quantify *F. virguliforme* detected from soybean roots using a ratio of PCR cycle thresholds for *F. virguliforme* and *G. max* gene targets (Wang et al., 2019). This study found no correlation between qPCR values and foliar symptoms. The authors speculate that this is a consequence of root resistance not being selected for in SDS-resistant varieties. This makes the QTL_17 an interesting candidate as a root resistance locus for further study. In addition to short-term benefits provided by resistance, root resistance also provides the long-term benefit of reducing *F. virguliforme* in the soil.

QTL_18a was identified in 5/6 population x year combinations and was tightly centered around the r*hg1/Rfs2* locus. It accounted for the largest proportion of phenotypic variance observed (36.4%) of any QTL identified in this study. This region has been identified as a major QTL by many SDS and SCN publications. The summary of SDS genetic mapping publications

identified 19 QTL mapping and GWAS publications for SDS resistance. Of these mapping studies, 10/19 reported this region on Chromosome Gm18 as important to SDS resistance (Chang et al., 1996; Chang et al., 1997; Njiti et al., 1998; Iqbal et al., 2001; Njiti et al., 2002; Kazi et al., 2008; Abdelmajid et al., 2012; Wen et al., 2014; Zhang et al., 2015; Tan et al., 2018). Most of these studies were conducted in a naturally infected SDS field. From the literature, *rhg1/Rfs2* is critical to SDS resistance in soybean.

The Decatur, MI field site used for phenotypic evaluations in this study is also used to evaluate SCN resistance, as this site has high levels of SCN. This study failed to collect SCN samples from field plots to assess SCN resistance, so connections between SDS and SCN resistance in this study are speculative. However, as was true in this study, it is interesting that the major SDS QTL identified in this study overlaps with the most widely studied SCN resistance gene, *rhg1*. This may be explained due to close proximities of the *rhg1* and *Rfs2* loci, as some evidence supports these loci as distinct, but closely linked (Meksem et al., 2001). However, it is of note that both the SDS resistant parents used in this study also carry SCN resistance derived from PI 88788, for which *rhg1* alone is sufficient for resistance to most races of SCN. In another recent study designed to map SDS resistance, Tan et al., (2018) identified SDS resistance QTL which colocalized over two SCN resistance genes, *rhg1* and *Rhg4*. The SDS resistant parents used in that study also contain SCN resistance derived from ‘Peking’, a resistance which requires two genes, *rhg1* and *Rhg4*, to be effective. This is not definitive evidence that SCN resistance may provide SDS resistance because resistance genes are often clustered together, and it may be possible that there are SDS resistance genes in close proximity to both r*hg1* and *Rhg4*. Fine mapping of SDS resistance genes should dissect whether they are

distinct from SCN resistance genes. Further study of the biology of SCN and SDS will also be useful to determine whether SCN resistance genes contribute to SDS resistance.

A resequencing study (Patil et al., 2019) of 106 elite, landrace, and wild soybean accessions examined copy number, haplotype, and promoter variation at the *rhg1* and *Rhg4* loci (Patil et al., 2019). It was concluded that haplotype variation, CNV, and promoter variation all play a role at both loci in mediating SCN resistance. The accessions from the resequencing study were grouped into six SCN resistance categories based on genomic variation at the *rhg1* and *Rhg4* loci. It may be interesting to evaluate lines from this study in an SCN and SDS infested field to further elucidate the relationship between SCN and SDS resistance. Interestingly, PI 407729 showed a broad resistance to SCN without a resistant genotype at *rhg1*. PI 407729 is susceptible to SDS based on field evaluations (Mueller et al., 2002). This seems to reinforce an independence in resistance mechanisms to SCN and SDS. PI 407729 may be an interesting candidate for studying the SDS and SCN resistance relationship.

In addition to identifying QTLs associated with SDS resistance, this study demonstrated the utility of using a spatial adjustment tool (R package MvngGrd) (Technow, 2011). This package has been used to adjust phenotypic data in an experiment examining barley resistance to stem rust (Nice et al., 2017). The adjustment pattern we used increased heritability from 14.3 to 15.6% for each trait, averaging a 15.0% increase. While comparisons of QTL mapping results were not explicitly reported in this study, anecdotal evidence suggests that the spatial data adjustment resulted in more narrow QTL peaks, corresponding to a smaller region of the genome being identified. While, the spatial adjustments increased heritability and thus increased QTL detection power, the data may be less easy to interpret. Without spatial adjustment, the additive effect of a QTL can serve as an estimate of the value of a phenotypic effect due to a resistance

allele. With adjusted phenotypic data, the additive effect is more difficult to interpret. However, the r^2 metric can still be used to estimate the proportion of phenotypic variance explained by the QTL, which can serve as a relative measure of the phenotypic effect of a given QTL.

LITERATURE CITED

LITERATURE CITED

Abdelmajid, K. M., Ramos, L., Leandro, L., Mbofung, G., Hyten, D. L., Kantartzi, S. K., et al. (2012). The 'PI 438489B' by 'Hamilton' SNP-based genetic linkage map of soybean [*Glycine max* (L.) Merr.] identified quantitative trait loci that underlie seedling SDS resistance. Journal of Plant Genome Sciences 1, 18-30.

Allen, T. W., Bradley, C. A., Sisson, A. J., Byamukama, E., Chilvers, M. I., Coker, C. M., et al. (2017). Soybean yield loss estimates due to diseases in the United States and Ontario, Canada, from 2010 to 2014. Plant Health Progress, 18, 19-27.

Anderson, J., Akond, M., Kassem, M. A., Meksem, K., and Kantartzi, S. K. (2015). Quantitative trait loci underlying resistance to sudden death syndrome (SDS) in MD96-5722 by 'Spencer' recombinant inbred line population of soybean. Biotech, 5, 203-210.

Aoki, T., O'Donnell, K., and Scandiani, M. M. (2005). Sudden death syndrome of soybean in South America is caused by four species of *Fusarium*: *Fusarium brasiliense* sp. nov., *F. cuneirostrum* sp. nov., *F. tucumaniae*, and *F. virguliforme*. Mycoscience, 46, 162-183.

Aoki, T., Scandiani, M. M., and O'Donnell, K. (2012). Phenotypic, molecular phylogenetic, and pathogenetic characterization of *Fusarium crassistipitatum* sp. nov., a novel soybean sudden death syndrome pathogen from Argentina and Brazil. Mycoscience, 53, 167-186.

Bao, Y., Kurle, J. E., Anderson, G., and Young, N. D. (2015). Association mapping and genomic prediction for resistance to sudden death syndrome in early maturing soybean germplasm. Molecular Breeding, 35, 128.

Brar, H. K., Swaminathan, S., and Bhattacharyya, M. K. (2011). The *Fusarium virguliforme* toxin FvTox1 causes foliar sudden death syndrome-like symptoms in soybean. Molecular Plant-Microbe Interactions, 24, 1179-1188.

Chang, S. J. C., Doubler, T. W., Kilo, V., Suttner, R., Klein, J., Schmidt, M. E., et al. (1996). Two additional loci underlying durable field resistance to soybean sudden death syndrome (SDS). Crop Science, 36, 1684-1688.

Chang, S. J. C., Doubler, T. W., Kilo, V. Y., Abu-Thredeih, J., Prabhu, R., Freire, V., et al. (1997). Association of loci underlying field resistance to soybean sudden death syndrome (SDS) and cyst nematode (SCN) race 3. Crop Science, 37, 965-971.

Chang, H. X., Roth, M. G., Wang, D., Cianzio, S. R., Lightfoot, D. A., Hartman, G. L., et al. (2018). Integration of sudden death syndrome resistance loci in the soybean genome. Theoretical and Applied Genetics, 131, 757-773.

Collins, P. J., Wen, Z., and Zhang, S. (2018). Marker-assisted breeding for disease resistance in crop plants. in Biotechnologies of Crop Improvement, Volume 3 (Springer, Cham.) 41-57.

Hartman, G. L., Chang, H. X., and Leandro, L. F. (2015). Research advances and management of soybean sudden death syndrome. Crop Protection, 73, 60-66.

Iqbal, M. J., Meksem, K., Njiti, V. N., Kassem, M. A., and Lightfoot, D. A. (2001). Microsatellite markers identify three additional quantitative trait loci for resistance to soybean sudden-death syndrome (SDS) in Essex× Forrest RILs. Theoretical and Applied Genetics, 102, 187-192.

Kandel, Y. R., Wise, K. A., Bradley, C. A., Chilvers, M. I., Tenuta, A. U., and Mueller, D. S. (2016). Fungicide and cultivar effects on sudden death syndrome and yield of soybean. Plant Disease, 100, 1339-1350.

Kandel, Y. R., Wise, K. A., Bradley, C. A., Chilvers, M. I., Byrne, A. M., Tenuta, A. U., et al. (2017). Effect of soybean cyst nematode resistance source and seed treatment on population densities of *Heterodera glycines*, sudden death syndrome, and yield of soybean. Plant Disease, 101, 2137-2143.

Kazi, S., Shultz, J., Afzal, J., Johnson, J., Njiti, V. N., and Lightfoot, D. A. (2008). Separate loci underlie resistance to root infection and leaf scorch during soybean sudden death syndrome. Theoretical and Applied Genetics, 116, 967-977.

Leandro, L. F. S., Eggenberger, S., Chen, C., Williams, J., Beattie, G. A., and Liebman, M. (2018). Cropping system diversification reduces severity and incidence of soybean sudden death syndrome caused by *Fusarium virguliforme*. Plant Disease, 102, 1748-1758.

Lightfoot, D. A. (2015). Two decades of molecular marker-assisted breeding for resistance to soybean sudden death syndrome. Crop Science, 55, 1460-1484.

Meksem, K., Pantazopoulos, P., Njiti, V. N., Hyten, L. D., Arelli, P. R., and Lightfoot, D. A. (2001). 'Forrest' resistance to the soybean cyst nematode is bigenic: saturation mapping of the *Rhg1* and *Rhg4* loci. Theoretical and Applied Genetics, 103, 710-717.

Mueller, D. S., Hartman, G. L., Nelson, R. L., and Pedersen, W. L. (2002). Evaluation of *Glycine max* germplasm for resistance to *Fusarium solani* f. sp. *glycines*. Plant Disease, 86, 741-746.

Nice, L. M., Steffenson, B. J., Blake, T. K., Horsley, R. D., Smith, K. P., and Muehlbauer, G. J. (2017). Mapping agronomic traits in a wild barley advanced backcross–nested association mapping population. Crop Science, 57, 1199-1210.

Njiti, V. N., Doubler, T. W., Suttner, R. J., Gray, L. E., Gibson, P. T., and Lightfoot, D. A. (1998). Resistance to soybean sudden death syndrome and root colonization by *Fusarium solani* f. sp. *glycine* in near-isogenic lines. Crop Science, 38, 472-477.

Njiti, V. N., Meksem, K., Iqbal, M. J., Johnson, J. E., Kassem, M. A., Zobrist, K. F., et al. (2002). Common loci underlie field resistance to soybean sudden death syndrome in Forrest, Pyramid, Essex, and Douglas. Theoretical and Applied Genetics, 104, 294-300.

Njiti, V. N., and Lightfoot, D. A. (2006). Genetic analysis infers Dt loci underlie resistance to *Fusarium solani* f. sp. *glycines* in indeterminate soybeans. Canadian Journal of Plant Science, 86, 83-90.

Nyquist, W. E., and Baker, R. J. (1991). Estimation of heritability and prediction of selection response in plant populations. Critical Reviews in Plant Sciences, 10, 235-322.

Patil, G. B., Lakhssassi, N., Wan, J., Song, L., Zhou, Z., Klepadlo, M., et al. (2019). Whole-genome re-sequencing reveals the impact of the interaction of copy number variants of the *rhg1* and *Rhg4* genes on broad-based resistance to soybean cyst nematode. Plant Biotechnology Journal, 17, 1595–1611.

Srour, A., Afzal, A. J., Blahut-Beatty, L., Hemmati, N., Simmonds, D. H., Li, W., et al. (2012). The receptor like kinase at *Rhg1-a/Rfs2* caused pleiotropic resistance to sudden death syndrome and soybean cyst nematode as a transgene by altering signaling responses. BMC Genomics, 13, 368.

Swaminathan, S., Abeysekara, N. S., Liu, M., Cianzio, S. R., and Bhattacharyya, M. K. (2016). Quantitative trait loci underlying host responses of soybean to *Fusarium virguliforme* toxins that cause foliar sudden death syndrome. Theoretical and Applied Genetics, 129, 495-506.

Tan, R., Serven, B., Collins, P. J., Zhang, Z., Wen, Z., Boyse, J. F., et al. (2018). QTL mapping and epistatic interaction analysis of field resistance to sudden death syndrome (*Fusarium virguliforme*) in soybean. Theoretical and Applied Genetics, 131, 1729-1740.

Tan, R., Collins, P. J., Wang, J., Wen, Z., Boyse, J. F., Laurenz, R. G., et al. (2019). Different loci associated with root and foliar resistance to sudden death syndrome (*Fusarium virguliforme*) in soybean. Theoretical and Applied Genetics, 132, 501-513.

Technow, F. (2011). R package mvngGrAd: Moving grid adjustment in plant breeding field trials. R package version 0.1, 2.

Tewoldemedhin, Y. T., Lamprecht, S. C., Geldenhuys, J. J., and Kloppers, F. J. (2014). First report of soybean sudden death syndrome caused by *Fusarium virguliforme* in South Africa. Plant Disease, 98, 569-569.

Tewoldemedhin, Y. T., Lamprecht, S. C., Vaughan, M. M., Doehring, G., and O'Donnell, K. (2017). Soybean SDS in South Africa is caused by *Fusarium brasiliense* and a novel undescribed *Fusarium* sp. Plant Disease, 101, 150-157.

Wang, J., Bradley, C. A., Stenzel, O., Pedersen, D. K., Reuter-Carlson, U., and Chilvers, M. I. (2017). Baseline sensitivity of *Fusarium virguliforme* to fluopyram fungicide. Plant Disease, 101, 576-582.

Wang, J., Sang, H., Jacobs, J. L., Oudman, K. A., Hanson, L. E., and Chilvers, M. I. (2019a). Soybean sudden death syndrome causal agent *Fusarium brasiliense* present in Michigan. Plant Disease, 103, 1234-1243.

Wang, J., Jacobs, J. L., Roth, M. G., and Chilvers, M. I. (2019b). Temporal dynamics of *Fusarium virguliforme* colonization of soybean roots. Plant Disease, 103, 19-27.

Wen, Z., Tan, R., Yuan, J., Bales, C., Du, W., Zhang, S., et al. (2014). Genome-wide association mapping of quantitative resistance to sudden death syndrome in soybean. BMC Genomics, 15, 809.

Zhang, J., Singh, A., Mueller, D. S., and Singh, A. K. (2015). Genome-wide association and epistasis studies unravel the genetic architecture of sudden death syndrome resistance in soybean. The Plant Journal, 84, 1124-1136.

CHAPTER 3

GENETIC MAPPING OF SOYBEAN OIL QUALITY TRAITS

I. Abstract

Soybean oil is predominantly composed of five fatty acids: palmitic acid, stearic acid, oleic acid, linoleic acid, and linolenic acid. Demand for high oleic acid soybean oil as a new food product is projected to grow rapidly in the future. Unlike conventional soybean oil, high oleic soybean oil does not require partial hydrogenation to have a high oxidative stability and long shelf life, and is therefore free of trans fats. The projected growth of this new product is due to many factors, including an FDA ban on trans fats in effect beginning in June 2018. High oleic soybean has been developed using several different methods by several different groups. Corteva and Bayer-Monsanto have used transgenic RNAi approaches to develop GMO high oleic soybeans, Plenish® and Vistive® Gold, respectively. Calytx™, a publicly traded biotech startup, has released a gene-edited high oleic soybean as their first commercial product. Many public soybean breeders are breeding non-GMO, high oleic soybean varieties using *GmFAD2*-1 mutants developed through mutagenesis. The high oleic acid trait can be combined with low linolenic acid content and low saturated fat content. The following study focused on a recombinant inbred line (RIL) population developed by crossing a high oleic, low linolenic, and low saturated fat content public variety (E16831) with a high yielding line with normal fatty acid content (E12076T). The parents, the RIL population, and bulked pools representing extreme phenotypes were genotyped with the SoySNP6K Illumina Infinium BeadChip, a linkage map was developed using JoinMap (v.2), and QTLs were detected using WinQTLCartographer (v 2.5). The study confirmed the effects of many known fatty acid biosynthesis genes including *GmFAD2-1B*, *GmFAD3A*, *GmFAD3B*, *GMFAD3C*, and *GmFATB-1A*. Possible strategies and limitations of

marker assisted selection (MAS) for the development of soybean varieties with high oleic acid, low linolenic acid, and low saturated fat content are discussed.

II. Introduction

Soybean (*Glycine max* (L.) *Merr.*) accounts for majority of the world's oilseed production, producing protein-rich soybean meal for animal feed and soybean oil used for frying, processed foods, and industrial applications (Lee et al., 2007). Critical characteristics of soybean oil are high oxidative stability and shelf life. Until recently, these qualities were improved in soybean oil through partial hydrogenation. However, the Food and Drug Administration changed the "generally regarded as safe" status of partially hydrogenated oils in 2018 due to the presence of trans-fats. This creates an urgent risk of soybean oil market share losses. For many decades, soybean researchers have studied rare soybean mutants with abnormal fatty acid profiles (Fehr, 2007). Some of these mutants have reduced levels of linoleic acid and elevated levels of oleic acid (Figure 3.1). Oil from these high oleic soybeans do not require partial hydrogenation to achieve high oxidative stability and shelf life, and is therefore a solution to the ban on trans-fats. Additionally, the health attributes of soybean oil can be improved through the reduction of the content of saturated fats in soybean oil.

Soybeans are approximately 20% oil by weight. Five constituent fatty acids compose the oil present in soybean seeds: Palmitic acid (16:0) (11%), stearic acid (18:0) (4%), oleic acid (18:1) (25%), linoleic acid (18:2) (52%), and linolenic acid (18:3) (8%) (Fehr, 2007) (Figure 3.1). Palmitic acid and stearic acid are saturated fatty acids; oleic acid is a monounsaturated fatty acid; linoleic and linolenic acid are polyunsaturated fatty acids. Oil from most soybean varieties are predominantly composed of linoleic acid, however the combination of mutated alleles of *GmFAD2-1A* (Wm82.a2.v1 Glyma.10g278000) and *GmFAD2-1B* (Wm82.a2.v1

Glyma.20g111000) changes the oleic acid content from ~20% to >75% of the total oil content (Alt et al., 2005; Pham et al., 2012; Kulkarni et al., 2017; Sweeney et al., 2017). Soybean varieties with oleic acid content >75% are referred to as high oleic soybeans. Mutated alleles of *GmFAD3A*, (Wm82.a2.v1 Glyma.14g194300), *GmFAD3B* (Wm82.a2.v1 Glyma.02g227200), and *GmFAD3C* (Wm82.a2.v1 Glyma.18g062000) have been demonstrated to reduce linolenic acid content from ~8% to less than 1% (Ross et al., 2000; Bilyeu et al., 2006; Bilyeu et al., 2011). Mutated alleles of *GmFATB-1A* (Wm82.a2.v1 Glyma.05g012300) and *GmKASIIIA* (Wm82.a2.v1 Glyma.09g277400) have been demonstrated to reduce saturated fats from ~15% to less than 8%, specifically via the reduction of palmitic acid content from ~11% to ~4% (Fehr et al., 1991; Burton et al., 1994; Primomo et al., 2002; Cardinal et al., 2007; De Vries et al., 2011; Cardinal et al., 2014; Gillman et al., 2014).

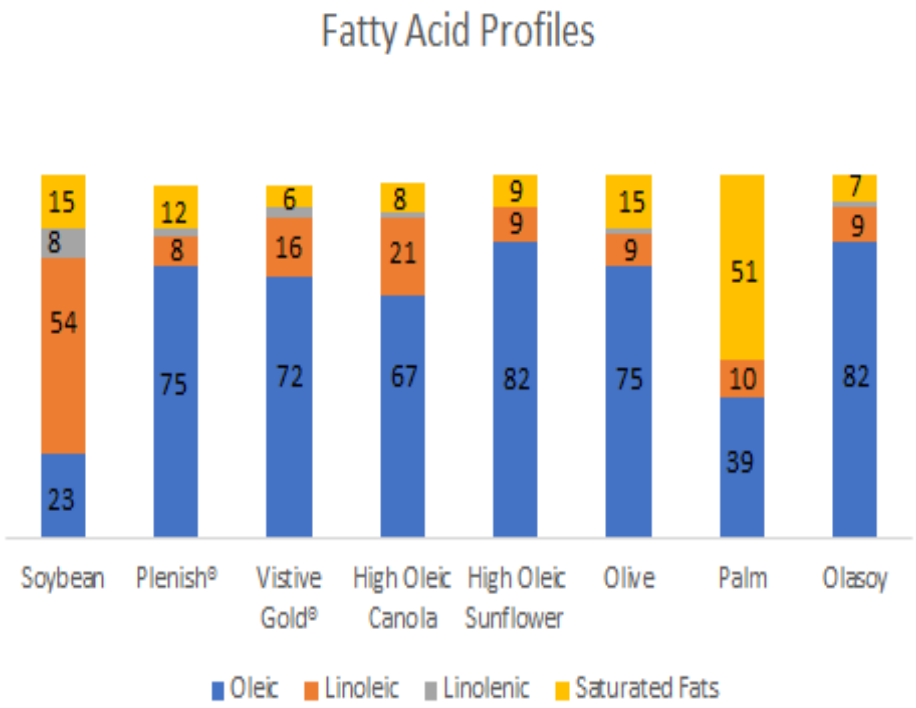


Figure 3.1. *Fatty acid profiles of common edible oils.* Common fry oils compared based on their composition of constituent fatty acids. Saturated fats are the sum of palmitic acid and stearic acid. Plenish® is a high oleic soybean from Corteva, Vistive Gold® is a high-oleic soybean from Bayer-Monsanto, and Olasoy is a high oleic soybean from Michigan State University Soybean Breeding Program

FAD2 genes code for enzymes critical to fatty acid biosynthesis (Figure 3.2). Specifically, they are responsible for the conversion of oleic acid into linoleic acid via the addition of a double bond at the 12th position on the fatty acid chain (Okuley et al., 1994). When the two *FAD2* genes are expressed in seed tissues, they are responsible for converting oleic acid to linoleic acid in developing soybean seeds, *GmFAD2-1A* and *GmFAD2-1B*. While silencing either *GmFAD2-1A* or *GmFAD2-1B* increases oleic content from 25% to about 40%, silencing both genes together greatly increases oleic content to over 80% of the oil content in the soybean seed. These genes have been silenced via several different strategies, but they all result in non-functional *GmFAD2* enzymes which can not convert oleic acid to linoleic acid, resulting in the

accumulation of oleic acid. High oleic soybeans were developed by DuPont scientists using a transgenic *GmFAD2-1* gene silencing construct (Kinney, 1995). The high oleic phenotype was also achieved using transgenic gene silencing of mRNA transcripts of *GmFAD2* genes via ribozyme termination (Buhr et al., 2002). Gene editing, using both TALENs (Haun et al., 2014; Demorest et al., 2016) and CRIPSR/Cas9 (Do et al., 2019), has been used to target *GmFAD2* genes of soybean to achieve a high oleic phenotype. A mutagenesis strategy has also been used to achieve a high oleic phenotype. See Chapter 1 for details of all the studies that have identified or induced mutations at *GmFAD2-1A* and *GmFAD2-1B*.

FAD3 genes code for enzymes also involved in fatty acid biosynthesis (Shah and Xin, 1997) (Figure 3.2). *FAD3* enzymes convert linoleic acid to linolenic acid by the addition of a third double bond to the fatty acid carbon chain. Studies have identified three *GmFAD3* genes in soybean which play a role in converting linoleic acid to linolenic acid in soybean seeds named *GmFAD3A*, *GmFAD3B*, and *GmFAD3C* (Pham et al., 2014). Studies have demonstrated that silencing *GmFAD3* genes decreases linolenic acid (Flores et al., 2008). Silencing *GmFAD3A* and either *GmFAD3B* or *GmFAD3C* reduces linolenic acid levels to below 3%, while silencing all three *GmFAD3* genes in soybean can reduce linolenic content to below 1% (Bilyeu et al., 2018). As with *GmFAD2* genes in soybean, silencing *GmFAD3* genes in soybean has been achieved through RNAi, gene editing, and mutagenesis (Flores et al., 2008; Demorest et al., 2016). See Chapter 1 for details on *GmFAD3* mutants identified and developed in soybean.

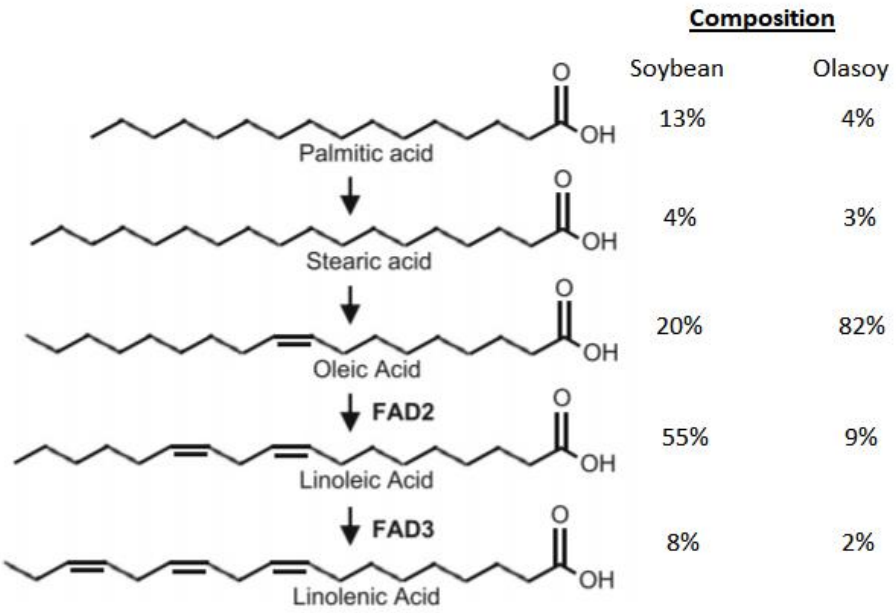


Figure 3.2. *Fatty acid biosynthesis genes in soybean.* The oil biosynthesis pathway and relevant genes for oil quality. Silencing *FAD2* and *FAD3* genes result in the accumulation of oleic acid and the reduction of linolenic acid.

Silencing *GmFAD2* and *GmFAD3* genes in soybean helps researchers increase the oil shelf life by shifting the oil profile to high oleic acid content and low linolenic acid content, as monounsaturated acids are more resistant to oxidation than polyunsaturated oils. Saturated fats are also resistant to oxidation, so increasing saturated fats could be another strategy for increasing oil shelf life. However, saturated fats are known to be unhealthy to humans, as high saturated fat diets can result in increased LDL cholesterol in the blood, leading to poor circulation, and elevated risk of cardiac arrest (Li et al., 2015). The ideal oil to balance shelf life and human heart health would get the increased shelf life benefits from high oleic and low linolenic, and heart health benefits from reduced saturated fats. Researchers have identified and developed mutant alleles at the *GmFATB-1A* and *GmKASIIIA* genes which reduce the palmitic content in soybean seeds (Cardinal et al., 2014; Gillman et al., 2014). Combining both of these

mutant alleles decreases total saturated fat content to less than 8%, which approximately corresponds to recommendations from the American Heart Association that people consume 1g of saturated fat or less per 14g serving of fat.

This study examined a recombinant inbred line population developed at MSU by crossing an MSU soybean variety with high oleic, low linolenic, and low saturated fats (E16831) with an MSU soybean variety with good agronomic traits, high yield, but lacking oil quality traits (E12076T). The RIL population segregated for all oil quality traits. A bulk segregant analysis, and a QTL mapping study was conducted on the RIL population to identify many of the genes that are known to be involved in the fatty acid biosynthesis pathway, as well as a modifier QTL related to linolenic acid content.

III. Methods

i. Population Development

The Michigan State University Soybean Breeding Program has developed soybean varieties with oleic acid > 75%, linolenic acid < 3%, and saturated fats > 8%. E16831 is a soybean variety developed by MSU which has 77% oleic acid, 1.6% linolenic acid, and 8.4% saturated fats. The pedigree of E16831 includes oil quality mutant lines developed by Walter Fehr at Iowa State, and Kristin Bilyeu at USDA-ARS in Missouri. E12076T is an elite conventional soybean variety developed by MSU which has superior yield and agronomic qualities, but has an oil quality profile similar to commodity soybeans (24% oleic acid, 7.7% linolenic acid, and 12% saturated fats). These lines were crossed, advanced to F4 by single seed descent, bulk harvested, and advanced to F4:6 RILs.

ii. Phenotyping

Two seeds were selected from each F4:6 line, crushed with a hydraulic press and custom metal plates, and oil was extracted using a hexane-based extraction. Oil samples were analyzed with a gas chromatograph (GC) and relative levels of five fatty acids were inferred from the area under the curve of peaks corresponding to palmitic acid, stearic acid, oleic acid, linoleic acid, and linolenic acid. Three biological replicates were collected for each RIL, and constituent fatty acid levels were averaged across replicates. From the 529 RILs, 118 RILs were selected based on uniformity between replicates, and selected to have normal distributions of oleic acid, linolenic acid, and saturated fats.

iii. Genotyping and linkage map construction

For the 118 RILs and the parents, young leaf tissue from ten plants from each F4:6 line was bulked and DNA was extracted using a modified CTAB extraction method. DNA concentration was diluted to 50 nanograms/microliter. The parents and the 118 RILs were genotyped with the SoySNP6K Illumina Infinium BeadChip. SNP calls from the SoySNP6k Illumina Infinium BeadChip was interpreted in Genome Studio.

A linkage map was constructed for the RIL population. SNP data was filtered to remove monomorphic markers based on the parental genotypes. SNP markers were filtered out if parental genotypes were missing, heterozygous, or shared between the parents. After filtering, 1969 markers were imported in JoinMap (ver. 2) and used to construct a linkage map specific to this population. A LOD of three was used to group markers and the regression algorithm was used to determine marker orders. Linkage groups with less than five markers were not included in the linkage maps. The linkage maps largely corresponded with the physical positions, but

some chromosomes were represented by two distinct linkage groups corresponding to different chromosome arms.

iv. Quantitative trait loci (QTL) mapping

The linkage map was imported into Windows QTL Cartographer (ver. 2.5) along with the mean of three replications of oil quality data (oleic acid, linolenic acid, and saturated fats). Composite interval mapping (CIM) was conducted using a walk speed of 0.5cM, forward and backward selection of markers as cofactors, and the Zmapqtl 6 model. The LOD threshold for QTL significance was determined for each population using a 1000 permutation test with alpha = 0.05. Flanking markers for each QTL were identified using a +/- 1LOD range from the LOD of the QTL peak.

v. Bulk segregant analysis (BSA)

In addition to genotyping 118 RILs for the QTL mapping analysis, six bulk pools of RILs were genotyped on the SoySNP6K Illumina Infinium BeadChip. The six bulk pools were each composed of six different RILs selected for uniformity of fatty acid contents between replicates, and selected to correspond to one of the six pools: high oleic, low oleic, high linolenic, low linolenic, high saturated fats, and low saturated fats. DNA from each of the six RILs corresponding to a given pool was combined prior to genotyping. Data was analyzed to identify genomic regions which were homozygous within a given bulk, shared by the parent with the corresponding oil trait, and opposite SNP calls between high and low bulks.

vi. Genotypic class analysis

Markers on the SoySNP6k Illumina Infinium BeadChip were selected to be adjacent to genes known to be involved in oil quality, specifically *GmFAD2-1B, GmFAD3A, GmFAD3B, GmFAD3C, GmFATB-1A*, and *GmKASIIIA*. Markers could not be identified in close proximity to *GmFAD2-1A*, as there were no polymorphic markers within 4.7 Mb of *GmFAD2-1A* between the RIL parents. The 118 RILs with genotypic data were classified into parental E16831 or E12076T genotypic classes for the known oil quality genes. One-way ANOVA tests for oleic acid content, linolenic acid content, and saturated fat content confirmed that genotypic classes had statistically significant effects on fatty acid contents. *Ad hoc*, two-tail T-tests with Bonferroni corrections were conducted to determine pairwise comparisons of genotypic classes to determine which classes had statistically significant differences between them.

For oleic acid content, lines were classified based on SNP genotypes flanking *GmFAD2-1B* as matching the normal oleic parent (E12076T) or the high oleic parent (E16831) (Figure 3.4a.). For the linolenic acid content, lines were classified based on SNP genotypes flanking *GmFAD3A, GmFAD3B*, and *GmFAD3C* as matching the normal linolenic parent (E12076T) or the low linolenic parent (E16831) (Figure 3.4b.). For saturated fat content, lines were classified based on SNP genotypes flanking *GmFATB-1A* and *GmKASIIIA* as matching the normal saturated fat content soybean parent (E12076T), or the low saturated fat content soybean parent (E16831) (Figure 3.4c.). If SNP markers flanking the above genes did not match, were missing, or were heterozygous, they were excluded from these comparisons. Including heterozygous markers in the genotypic classification system, would allow comparisons to dissect dominance effects of the above genes, but additive effects are more important in breeding oil quality traits into a crop which is planted as inbred lines, such as soybean.

III. Results

i. Phenotyping

For all 529 RILs and both parents, three replicates of gas chromatograph (GC) fatty acid data were collected and averaged. While GC data included all five fatty acids separately, data was analyzed focusing on oleic, linolenic, and saturated fat content (palmitic acid + stearic acid). As to be expected with an F4-derived RIL population, some RILs were segregating for oil quality traits as observed by large variability between replicates. Lines with large variability between replicates were not selected for genotyping, as homozygous individuals are more informative for inferring additive effect, which is more relevant in breeding inbred crops. For the three traits analyzed, the populations largely followed a normal distribution (Figure 3.3). Transgressive segregation was observed for all traits. Oleic acid content ranged from 20.87% to 82.27%, linolenic acid content ranged from 0.94% to 11.01%, and saturated fat content ranged from 6.59% to 14.41% (Table 3.1).

Table 3.1. Fatty acid contents*. Fatty acid contents for RIL and parents were calculated based on gas chromatography. Data is presented as a percentage composition of consistent fatty acids.*

	Population Mean (%)	**Range (%)**	**E12076T (%)**	**E16831 (%)**
Oleic	47.05	20.87 - 82.27	24.40	76.55
Linolenic	4.61	0.94 - 11.01	7.77	1.61
Saturated Fats	10.30	6.59 - 14.41	12.01	8.42

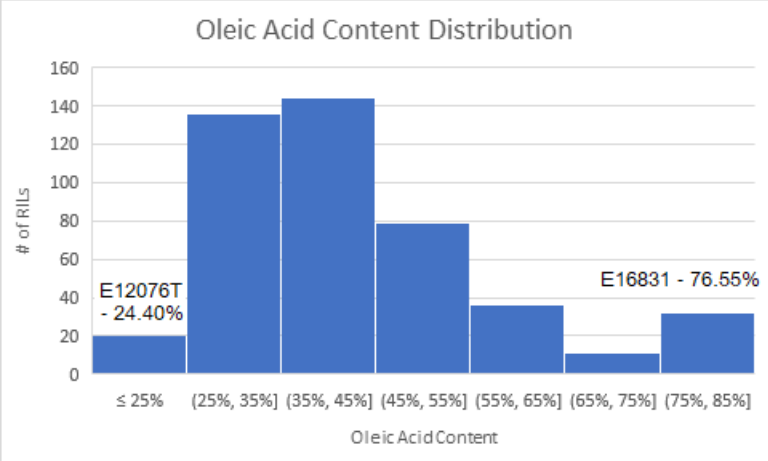


Figure 3.3a. *Phenotypic distribution of oleic acid content.* RIL distribution of oleic acid content as calculated from gas chromatography. Parental oleic acid content is listed within the figure.

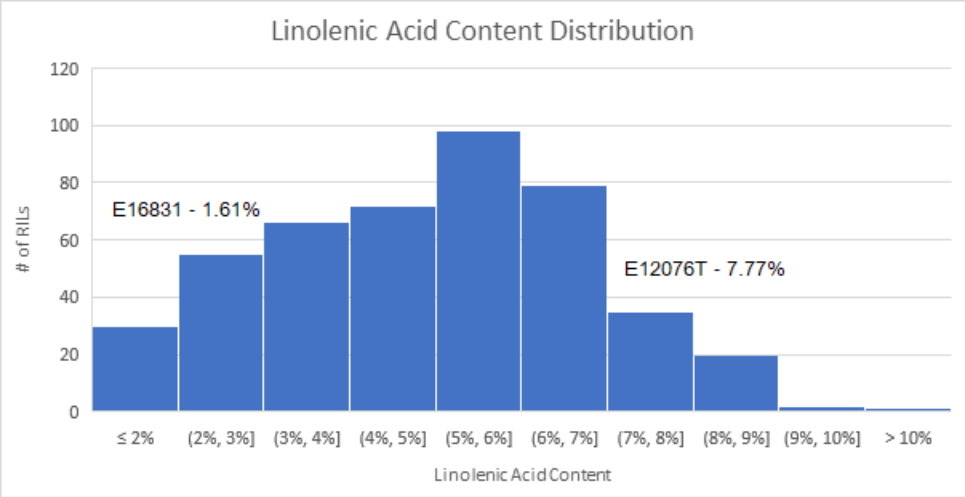


Figure 3.3b. *Phenotypic distribution of linolenic acid content.* RIL distribution of linolenic acid content as calculated from gas chromatography. Parental linolenic acid content is listed within the figure.

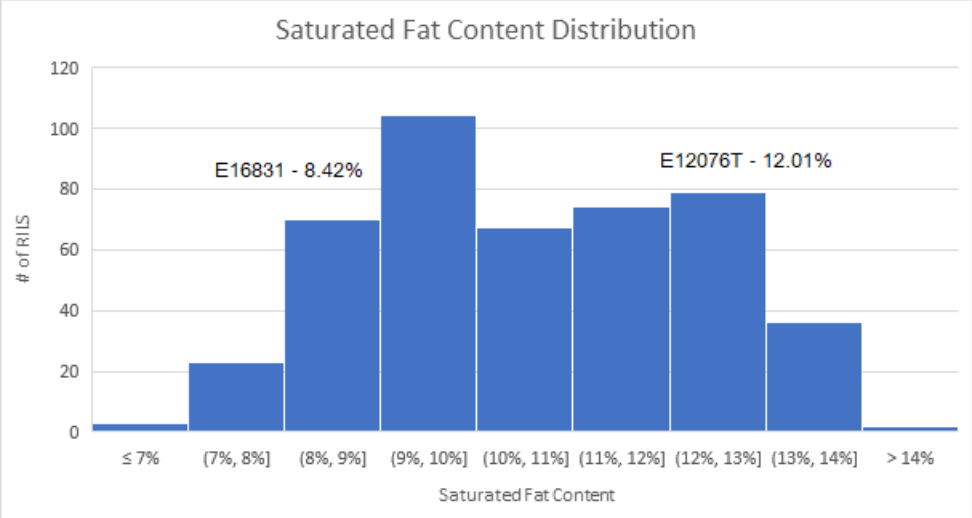


Figure 3.3c. *Phenotypic distribution of saturated fat content.* RIL distribution of saturated fat content as calculated from gas chromatography. Parental saturated fat content is listed within the figure.

ii. Heritability

Broad-sense heritability was calculated after running an ANOVA for each oil trait and using the mean square values representing genetic variance and environmental variance based on the following formula from Nyquist and Baker 1991: $h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2 / r}$. The heritability values are summarized in the Table 3.2. Heritability values were high for oleic acid content (0.93) and linolenic acid content (0.93). Saturated fat content heritability was moderate (0.49), so each constituent saturated fat heritability was also measured. Palmitic acid content had moderate to high heritability (0.78), confirming previous reports that a low saturated fat phenotype is best targeted by aiming to lower palmitic acid content.

Table 3.2 – Broad sense heritability*. Heritability was calculated for oleic acid, linolenic acid, and saturated fat content for the RIL population. 90% lower and upper confidence intervals were also calculated.*

	Heritability	**90% CI lower limit**	**90% CI upper limit**
Oleic Acid Content	0.93	0.92	0.94
Linolenic Acid Content	0.93	0.92	0.94
Saturated Fat Content	0.49	0.42	0.55
Palmitic Acid Content	0.78	0.76	0.81
Stearic Acid Content	0.55	0.49	0.61

iii. Genotyping and linkage mapping

A linkage map was constructed for the RIL population E16831 x E12076T using SoySNP6k Illumina Infinium BeadChip data from parent lines and 118 RILs (Table 3.3). After filtering based on parent genotypes, and segregation distortion, 1556 SNP markers were used to construct the map. JoinMap ver. 2 software was used to group markers in 26 linkage groups. Linkage groups were composed of 5 – 182 markers. A regression algorithm was used to order the markers within each linkage group. The size of linkage groups ranged from 4.64cM to 238.14cM. The average distance between markers was 1.81cM.

Table 3.3. Linkage map statistics*. The linkage map was calculated using JoinMap ver. 2.*

Population FA			**Linkage Groups**		**Chromosomes**	
Chromosome	Linkage Group	#Markers	Linkage Group Size (cM)	Average Distance (cM) between Markers	Chromosome Size (cM)	Average Distance (cM) between Markers
Gm01	LG01.1	5	4.64	0.93	102.97	2.34
	LG01.2	39	98.33	2.52		
Gm02	LG02	70	170.65	2.44	170.65	2.44
Gm03	LG03	75	188.34	2.51	188.34	2.51
Gm04	LG04.1	88	80.54	0.92	104.60	1.05
	LG04.2	12	24.06	2.00		
Gm05	LG05	112	153.83	1.37	153.83	1.37
Gm06	LG06.1	63	131.51	2.09	146.76	1.88
	LG06.2	15	15.25	1.02		
Gm07	LG07	63	191.69	3.04	191.69	3.04
Gm08	LG08.1	65	162.22	2.50	176.21	2.18
	LG08.2	16	13.99	0.87		
Gm09	LG09	85	163.80	1.93	163.80	1.93
Gm10	LG10	74	147.17	1.99	147.17	1.99
Gm11	LG11	34	156.98	4.62	156.98	4.62
Gm12	LG12	23	88.38	3.84	88.38	3.84
Gm13	LG13	182	238.14	1.31	238.14	1.31
Gm14	LG14	69	130.83	1.90	130.83	1.90
Gm15	LG15.1	18	63.70	3.54	76.97	2.75
	LG15.2	10	13.27	1.33		
Gm16	LG16	64	84.03	1.31	84.03	1.31
Gm17	LG17	109	155.62	1.43	155.62	1.43
Gm18	LG18	93	143.42	1.54	143.42	1.54
Gm19	LG19.1	57	45.71	0.80	80.96	0.83
	LG19.2	41	35.24	0.86		
Gm20	LG20	74	121.32	1.64	121.32	1.64
Total	**26**	**1556**	-	-	**2822.66**	**1.81**

iv. BSA results

From the six bulk pools, genotype data was compared between low and high bulk pool corresponding to the three oil traits (oleic acid, linolenic acid, and saturated fat contents).

Comparing the high oleic bulk to the low oleic bulk revealed associations of 2 or more markers on Chromosomes Gm04, Gm19, and Gm20. The association on Gm20 included the *GmFAD2-1B* gene, known to be critical for the high oleic phenotype. *GmFAD2-1A* was not detected, as there were zero polymorphic SNP markers within 4.7 Mb of the *GmFAD2-1A* gene. Comparing the low linolenic bulk with the high linolenic bulk revealed associations on Chromosomes Gm02, Gm14, Gm18, and Gm20. These associations included regions corresponding with all three *GmFAD3* genes, known to play roles in linolenic acid synthesis. Comparison of the low saturated fats bulk with the high saturated fats bulk revealed associations on Chromosomes Gm04, Gm05, Gm15, and Gm18. The associations on Gm05 were in close proximity to *GmFATB-1A*, of which mutant alleles have been shown to reduce palmitic, and thus, saturated fat content. *GmKASIIIA* was not detected by bulk segregant analysis.

v. QTL mapping

A total of five QTL were detected by the QTL analysis, three for linolenic content, and two for saturated fats (Table 3.4 and Figure 3.5). For linolenic content, QTLs were detected on Chromosomes Gm02, Gm14, and Gm19. For saturated fat content, QTLs were detected on Chromosomes Gm05 and Gm09.

Table 3.4. Summary of quantitative trait loci (QTL)*. QTL were detected using the linkage map and gas chromatography data.*

QTL	Trait	Chromosome	Peak Physical Position[a] (Mb)	Flanking Markers [a] (Mb)	Peak LOD	R^2	Beneficial Allele Derived from	Gene
QTL-02	Linolenic	Gm02	40.81	39.85 - 41.54	7.58	6.33%	E16831	GmFAD3B (41.41 Mb)
QTL-05	Saturated Fats	Gm05	1.09	0.31 - 2.07	21.63	53.12 %	E16831	GmFATB-1A (1.13 Mb)
QTL-09	Saturated Fats	Gm09	43.4	42.79 - 43.82	4.32	3.78%	E12076T	GmKASIIIA[b] (49.28 Mb)
QTL-14	Linolenic	Gm14	45.99	45.57 - 46.13	24.35	46.66 %	E16831	GmFAD3A (45.94 Mb)
QTL-19	Linolenic	Gm19	44.52	43.32 - 44.84	3.99	2.76%	E12076T	Hyten et al., 2004

[a] Physical positions are based on Wm82.a2.v1

[b] Note that GmKASIIIA is about 5 Mb outside of QTL-09

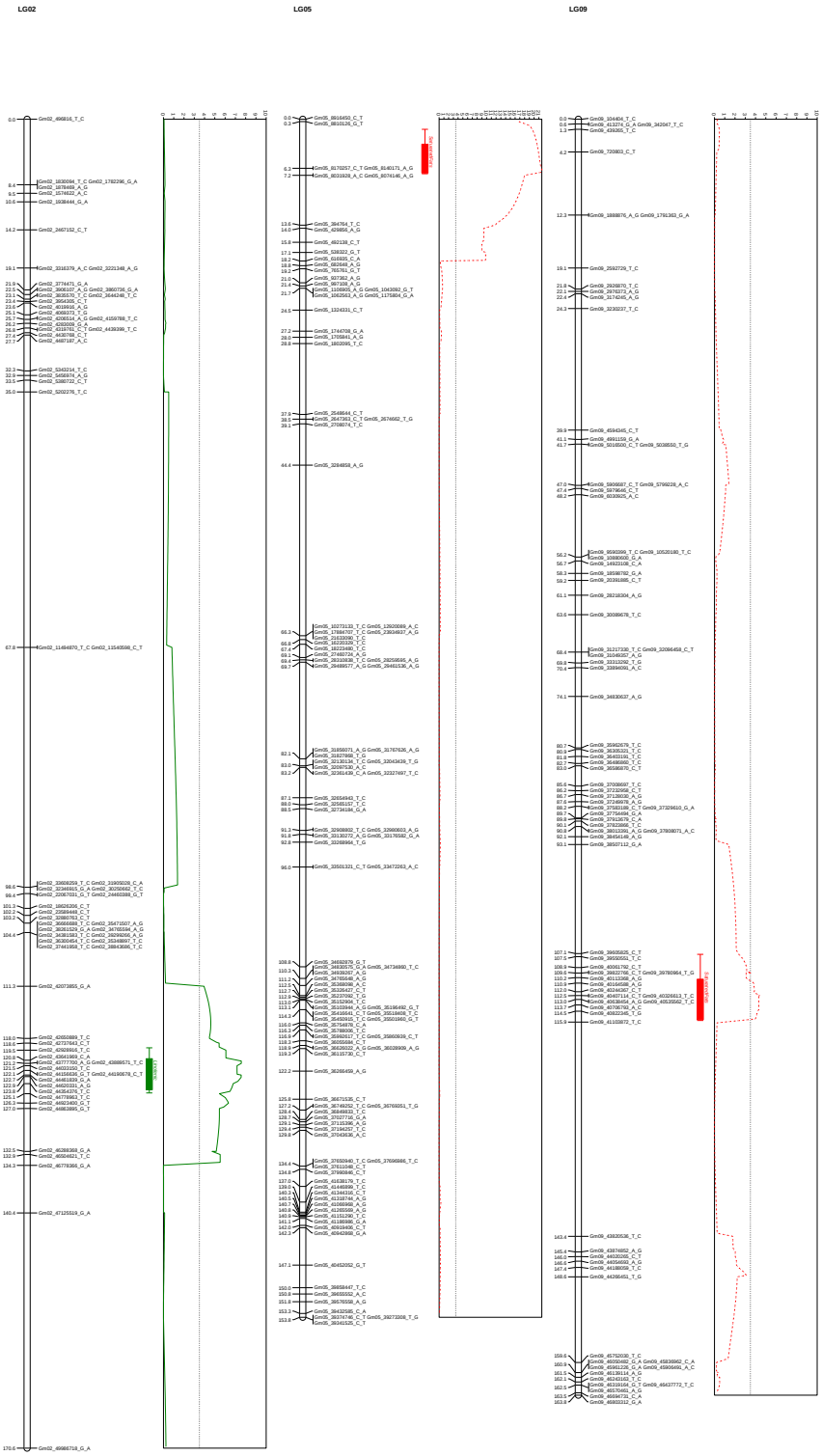
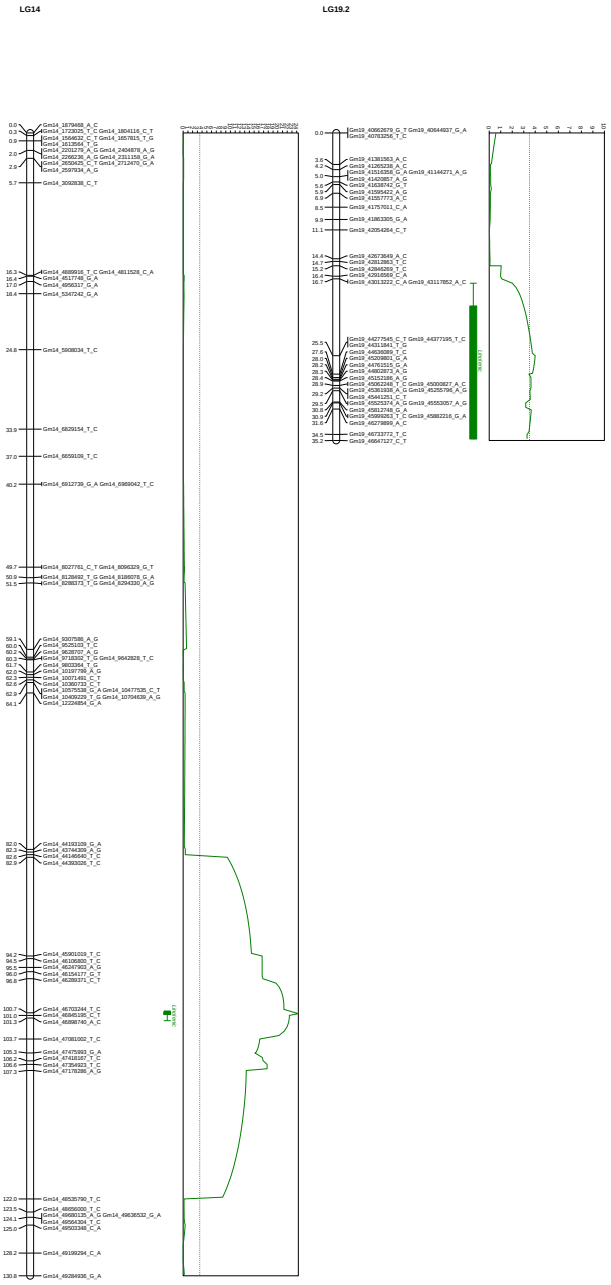


Figure 3.4. *QTL positions on linkage groups.* This is a figure of the linkage groups which contained significant QTLs for linolenic acid and saturated fat content. See **Table 3.4** for a summary of QTL statistics.

Figure 3.5. (cont'd)



QTL-02 was detected for linolenic acid content, had a LOD peak at physical position 40.81 Mb on Chromosome Gm02, and accounted for 6.33% of phenotypic variation observed. The range of QTL-02 includes *GmFAD3B*. QTL-05 was detected for saturated fat content, had a LOD peak at physical position 1.09 Mb on Chromosome Gm05, and accounted for 53.12% of

phenotypic variation observed. QTL-05 includes *GmFATB-1A*. QTL-09 was detected for saturated fat content, had a LOD peak at physical position 43.40 Mb on Chromosome Gm09, and accounted for 3.78% of phenotypic variation observed. QTL-09 was in proximity to *GmKASIIIA*, but the fatty acid gene was about 5 Mb from the QTL boundary. QTL-14 was detected for linolenic acid content, had a LOD peak corresponding to the physical position 45.99 Mb on Chromosome Gm14, and accounted for 46.66% of phenotypic variation observed. The QTL-14 region included the gene *GmFAD3A*. QTL-19 was detected for linolenic acid content, had a LOD peak at physical position 44.52 Mb on Chromosome Gm19, and accounted for 2.76% of phenotypic variation observed. The QTL-19 region overlapped with a QTL detected for linolenic acid content in a Williams X Essex RIL population (Hyten et al., 2004). The authors of that publication describe this QTL region as a "modifier QTL", as neither parent in their population contained low linolenic mutant alleles.

No QTLs were detected for oleic acid content with LOD scores greater than 3.0. It is of note that there were no polymorphic markers within 4.7Mb of *GmFAD2-1A*. *GmFAD2-1B* did have polymorphic markers in close vicinity, and was detected in the bulk-segregant analysis. However, the effects of mutant alleles *GmFAD2-1B* is most strongly observed in conjunction with mutant alleles of *GmFAD2-1A*. As has been well described in the literature, it is likely that *GmFAD2-1A* and *GmFAD2-1B* were responsible for the high oleic phenotype in this population, but marker limitations prevented their detection.

vi. Genotypic class analysis

For oleic acid content, the lack of polymorphism in *GmFAD2-1A* meant that the only classes compared were RILs with a E12076T-type *GmFAD2-1B* and a E16831-type *GmFAD2-1B*. The

two-tailed t-test comparison between these two groups was statistically significant at alpha=0.05 with the E12076T-type *GmFAD2-1B* class (n=32) averaging 37.50% oleic acid content and the E16831-type *GmFAD2-1B* class (n=15) averaging 48.27% oleic acid content (Figure 3.4).

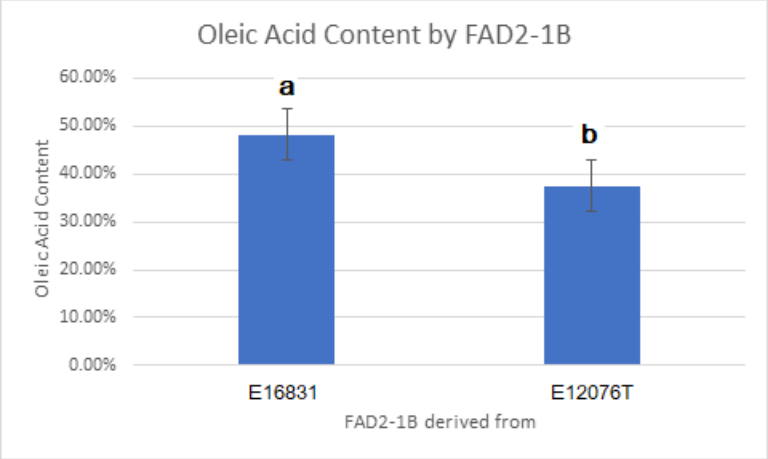


Figure 3.5a. *Genotypic classes and oleic acid contents.* Known genes involved in oleic acid content are *GmFAD2-1A* and *GmFAD2-1B*. Polymorphic markers were only detected for *GmFAD2-1B* in the RIL population. RILs were classified into genotypic classes corresponding to *GmFAD2-1B* derived from E16831 or *GmFAD2-1B* derived from E12076T. A t-test determined that the oleic acid content was significantly (alpha = 0.05) between these two classes, indicated by different letters over the bar graphs.

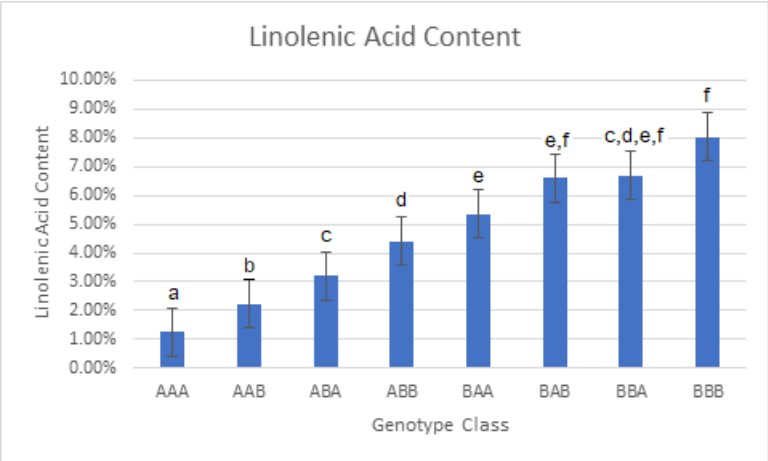


Figure 3.5b. *Genotypic classes and linolenic acid contents.* Known genes involved in linolenic acid content are *GmFAD3A*, *GmFAD3B*, and *GmFAD3C*. RILs were classified into genotypic classes corresponding to *GmFAD3A* derived from E16831 (A) or E12076T (B), *GmFAD3B*

derived from E16831 (A) or E12076T (B), and *GmFAD3C* derived from E16831 (A) or E12076T (B). Pairwise t-tests determined that the linolenic acid content was significantly different (Bonferroni corrected alpha = 0.05) between classes. Different letters indicate that genotypic classes are significantly different. This data suggests that *GmFAD3A* has a large, significant effect on linolenic content, while *GmFAD3B* and *GmFAD3C* have significant effect on linolenic content only when combined with the E16831-type *GmFAD3A*.

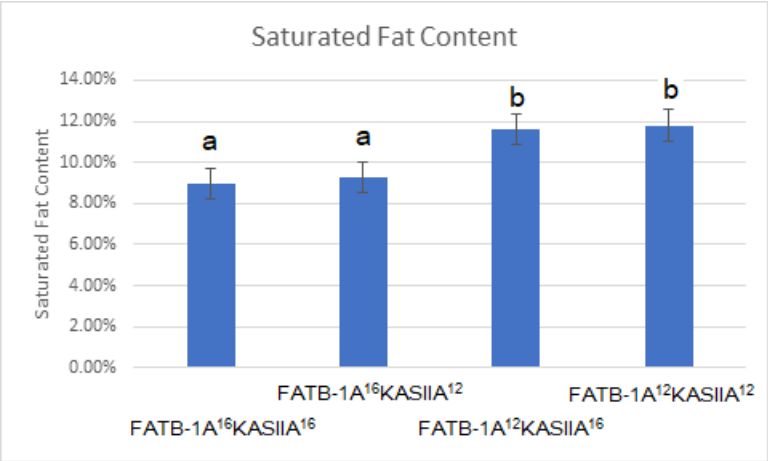


Figure 3.5c. *Genotypic classes and saturated fat contents.* Known genes involved in saturated fat content are *GmFATB-1A* and *GmKASIIIA.* RILs were classified into genotypic classes corresponding to *GmFATB-1A* derived from E16831 or E12076T and *GmKASIIIA* derived from E16831 or E12076T. Pairwise t-tests determined that the saturated fat content was significantly different (Bonferroni corrected alpha = 0.05) between classes based on *GmFATB-1A* genotype, but not different based on *GmKASIIIA* genotype. Different letters indicate that genotypic classes are significantly different.

For linolenic acid content, there were eight genotypic classes corresponding to genotype calls of E12076T or E16831 at *GmFAD3A*, *GmFAD3B*, and *GmFAD3C*. As an example, $A_{16}B_{16}C_{12}$ is an abbreviation which refers to the genotypic class with E16831-like genotypes at *GmFAD3A* and *GmFAD3B*, and with E12076T-like genotypes at *GmFAD3C*. $A_{16}B_{16}C_{16}$ (n=5) had a significantly lower linolenic acid content (1.26%) than all other genotypic classes. $A_{16}B_{16}C_{12}$ (n=6) was significantly higher in linolenic acid content (2.24%) compared with $A_{16}B_{16}C_{16}$, but lower than all other classes. $A_{16}B_{12}C_{16}$ (n=6) was significantly higher in linolenic acid content (3.19%) than $A_{16}B_{16}C_{16}$ and $A_{16}B_{16}C_{12}$, and significantly lower than all other

genotypic classes except for $A_{12}B_{12}C_{16}$. $A_{16}B_{12}C_{12}$ (n=5) was significantly higher in linolenic acid content (4.23%) compared with all other classes with the E16831-type *GmFAD3A,* but significantly lower than the other classes except for $A_{12}B_{12}C_{16}$. $A_{12}B_{16}C_{16}$ (n=8) was significantly higher in linolenic acid content (5.36%) compared with all classes with the E16831-type *GmFAD3A,* and significantly lower than $A_{12}B_{12}C_{12}$. $A_{12}B_{16}C_{12}$ (n=4) was significantly higher in linolenic acid content (6.61%) than all classes with the E16831-type *GmFAD3A*. $A_{12}B_{12}C_{16}$ (n=5) had the largest variance of any class, and was only significantly higher in linolenic acid content (6.69%) than $A_{16}B_{16}C_{16}$ and $A_{16}B_{16}C_{12}$. $A_{12}B_{12}C_{12}$ (n=6) had the lowest average linolenic content (8.02%) and was significantly lower than all classes with the E16831-type *GmFAD3A* and $A_{12}B_{16}C_{16}$. Based on the above comparisons, this analysis concludes that all three alleles *GmFAD3* genes derived from E16831 work to lower linolenic acid levels. *GmFAD3A* seems to have the largest and most consistent effect on linolenic acid levels, while *GmFAD3B* and *GmFAD3C* both influence linolenic acid levels only when combined with the E16831-type *GmFAD3A*.

For saturated fat content, there were four genotypic classes corresponding to the genotype calls of E12076T or E16831 at *GmFATB-1A* and *GmKASIIIA*. FATB-$1A_{16}$KASIIIA_{16} (n=12) and FATB-$1A_{16}$KASIIIA_{12} (n=15) had significantly lower saturated fat content (8.97% and 9.27%, respectively) than classes with the E12076T locus at the *GmFATB-1A* locus, but were not significantly different from each other. Likewise, FATB-$1A_{12}$KASIIIA_{16} (n=16) and FATB-$1A_{12}$KASIIIA_{12} (n=17) had significantly higher saturated fat content (11.61% and 11.81%, respectively) than classes with the E16831 locus at the *GmFATB-1A* locus, but were not significantly different from each other. The combined analysis reinforces the QTL mapping results that *GmFATB-1A* plays a significant role in saturated fat content in this population.

However, the analysis also fails to find evidence that *GmKASIIIA* is playing a role in saturated fat content in this population.

IV. Discussion

In summary, these genetic mapping analyses confirmed the effects of known fatty acid biosynthesis genes in a population segregating for three different oil quality traits. Bulk-segregant analysis successfully identified genomic regions associated with *GmFAD2-1B, GmFAD3A, GmFAD3B, GmFAD3C,* and *GmFATB-1A* based on genotyping high and low bulks corresponding to their fatty acid contents. Linkage map development and QTL mapping identified five QTLs corresponding to oil quality traits. Three of these QTLs were associated with linolenic acid content, and two of the QTLs are associated with saturated fat content. Four of these QTL appeared to overlap with known fatty acid biosynthesis genes (although evidence for linkage between QTL-09 and *GmKASIIIA* was tenuous in the QTL mapping analysis and not supported by bulk segregant analysis or genotypic class analysis), and one overlapped with a previously identified QTL detected in a different soybean RIL population (Hyten et al., 2004).

The QTL mapping analysis revealed several opportunities and challenges towards applying the well-studied genetic basis of fatty acid contents towards more efficient breeding of soybean varieties with a specific desired fatty acid profile. While we detected five QTL, only two QTL had R^2 values > 10%. QTL-05 accounted for 53.12% of the phenotypic variance observed for saturated fat content and overlapped with *GmFATB-1A*. QTL-14 accounted for 46.66% of the phenotypic variance observed for linolenic acid content and overlapped with *GmFAD3A*. QTL-05 and QTL-14 are major effect QTL for saturated fat content, and linolenic

acid, respectively, making them good candidates to explore using marker assisted selection (MAS) for these loci.

However, MAS for oil quality traits is unlikely to completely replace phenotypic selection for several reasons. One reason is related to the challenges this study had in identifying *GmFAD2-1A*, a gene known to be critical towards the high oleic phenotype. We failed to identify *GmFAD2-1A* in either the bulk-segregant analysis or the QTL mapping. The absence of polymorphic markers within 4.7 Mb of the genomic location of *GmFAD2-1A*. This is most likely related to the nature of how the high oleic trait was developed. Chapter 1 summarized the many sources of *GmFAD2-1A* mutations. Most of these sources were developed via chemical or X-ray mutagenesis of varieties 'Bay' and 'Williams 82', suggesting that the functional changes in the genomes of the mutant lines are due to point mutations or indels directly within the *GmFAD2-1A* gene (Sandhu et al., 2007; Anai et al., 2008; Thapa et al., 2016; Combs and Bilyeu, 2019). These small mutations have major effects on the oleic acid content in the soybean seeds, but there is little detectable change in the genomic regions immediately surrounding the gene of interest, making it difficult to identify polymorphic markers in this region unless they are directly linked to the causal mutation. While the SoySNP6K Illumina Infinium BeadChip is a convenient genotyping platform, it failed to identify any polymorphic markers linked to *GmFAD2-1A* in the RIL population used in this study. Based on pedigree information, and given the limited genetic diversity described in the North American elite soybean line gene pool, it would not be surprising if the genomic region surrounding *GmFAD2-1A* in E12076T was derived from a wild-type 'Williams 82' and the genomic region surrounding *GmFAD2-1A* in E16831 was derived from a mutant derived from 'Williams 82'. Unless a genetic marker identified the causative polymorphism differentiating the wild-type *GmFAD2-1A* from the mutant *GmFAD2-1A*, other

markers adjacent to the *GmFAD2-1A* gene would likely fail to differentiate E12076T haplotypes and E16831 haplotypes in our population. While this phenomenon was observed specifically for the *GmFAD2-1A* locus in this study, a similar phenomenon of no detectable polymorphisms near the causative genes could limit the applicability of MAS to be used in populations segregating for oil quality. Of course, 'perfect markers', which directly detect the causative mutation in oil quality genes, would overcome this problem, and many have been developed which detect the causative mutations in fatty acid biosynthesis genes (Anai et al., 2008; Pham et al., 2010).

The other major limitation in applying MAS for oil quality is that there are many minor genes and QTL which have small effects (R^2< 10%) at increasing or decreasing constituent fatty acid contents. MAS is most efficiently applied when traits are qualitatively controlled by a single or few genes, as the effects of minor genes and QTLs are often overlooked by MAS (Collins et al., 2018). This analysis suggests that a good approach for a soybean breeder would be to use markers linked to the large effect genes, *GmFATB-1A* and *GmFAD3A*, to cull breeding populations against wild-type alleles at these loci, followed by a phenotypic selection of breeding populations to identify the breeding lines that meet the oil quality targets. This approach would use MAS as a negative selection tool to reduce the number of breeding lines which need to have fatty acid content measured via GC, but would not replace phenotypic selection for oil quality traits.

This study also confirmed previous reports that stacking fatty acid mutations for high oleic, low linolenic, and low saturated fats can result in soybean lines with all three oil quality traits (Bilyeu et al., 2018). The target oil profile for the Michigan State University Soybean Breeding Program is high oleic acid content (>75%), low linolenic acid content (<3%), and low saturated fat content (<8%). These traits have been stacked into soybean varieties, patented under

the name ‘Olasoy’, whose oil will have a longer shelf life without hydrogenation, be better for cardiovascular health than soybean oil from commodity soybeans, and be marketable as ‘non-GMO products ’. Further studies are ongoing to test the hypothesis that oil from ‘Olasoy’ varieties will be more resistant to oxidation during frying than oil from commodity soybeans. This ongoing study will also assess the flavor of French fries cooked in oil from ‘Olasoy’ varieties to confirm that this oil has an acceptable flavor profile.

LITERATURE CITED

LITERATURE CITED

Alt, J. L., Fehr, W. R., Welke, G. A., and Shannon, J. G. (2005). Transgressive segregation for oleate content in three soybean populations. Crop Science, 45, 2005-2007.

Anai, T., Yamada, T., Hideshima, R., Kinoshita, T., Rahman, S. M., and Takagi, Y. (2008). Two high-oleic-acid soybean mutants, M23 and KK21, have disrupted microsomal omega-6 fatty acid desaturase, encoded by *GmFAD2-1A*. Breeding Science, 58, 447-452.

Bilyeu, K., Palavalli, L., Sleper, D. A., and Beuselinck, P. (2006). Molecular genetic resources for development of 1% linolenic acid soybeans. Crop Science, 46, 1913-1918.

Bilyeu, K., Gillman, J. D., and LeRoy, A. R. (2011). Novel *FAD3* mutant allele combinations produce soybeans containing 1% linolenic acid in the seed oil. Crop Science, 51, 259-264.

Bilyeu, K., Škrabišová, M., Allen, D., Rajcan, I., Palmquist, D. E., Gillen, A., et al. (2018). The interaction of the soybean seed high oleic acid oil trait with other fatty acid modifications. Journal of the American Oil Chemists' Society, 95, 39-49.

Buhr, T., Sato, S., Ebrahim, F., Xing, A., Zhou, Y., Mathiesen, M., et al. (2002). Ribozyme termination of RNA transcripts down-regulate seed fatty acid genes in transgenic soybean. The Plant Journal, 30, 155-163.

Burton, J. W., Wilson, R. F., and Brim, C. A. (1994). Registration of N79-2077-12 and N87-2122-4, two soybean germplasm lines with reduced palmitic acid in seed oil. Crop Science, 34, 313.

Cardinal, A. J., Burton, J. W., Camacho-Roger, A. M., Yang, J. H., Wilson, R. F., and Dewey, R. E. (2007). Molecular analysis of soybean lines with low palmitic acid content in the seed oil. Crop Science, 47, 304-310.

Cardinal, A. J., Whetten, R., Wang, S., Auclair, J., Hyten, D., Cregan, P., et al. (2014). Mapping the low palmitate *fap1* mutation and validation of its effects in soybean oil and agronomic traits in three soybean populations. Theoretical and Applied Genetics, 127, 97-111.

Collins, P. J., Wen, Z., and Zhang, S. (2018). Marker-assisted breeding for disease resistance in crop plants. in Biotechnologies of Crop Improvement, Volume 3 (Springer, Cham.) 41-57.

Combs, R., and Bilyeu, K. (2019). Novel alleles of *FAD2-1A* induce high levels of oleic acid in soybean oil. Molecular Breeding, 39, 79.

Demorest, Z. L., Coffman, A., Baltes, N. J., Stoddard, T. J., Clasen, B. M., Luo, S., et al. (2016). Direct stacking of sequence-specific nuclease-induced mutations to produce high oleic and low linolenic soybean oil. BMC Plant Biology, 16, 225.

De Vries, B. D., Fehr, W. R., Welke, G. A., and Dewey, R. E. (2011). Molecular characterization of the mutant *fap3* (A22) allele for reduced palmitate concentration in soybean. Crop Science, 51, 1611-1616.

Do, P. T., Nguyen, C. X., Bui, H. T., Tran, L. T., Stacey, G., Gillman, J. D., et al. (2019). Demonstration of highly efficient dual gRNA CRISPR/Cas9 editing of the homeologous *GmFAD2–1A* and *GmFAD2–1B* genes to yield a high oleic, low linoleic and α-linolenic acid phenotype in soybean. BMC Plant Biology, 19, 311.

Fehr, W. R. (2007). Breeding for modified fatty acid composition in soybean. Crop Science, 47(Supplement_3), S-72.

Fehr, W. R., Welke, G. A., Cianzio, S. R., Duvick, D. N., and Hammond, E. G. (1991). Inheritance of reduced palmitic acid content in seed oil of soybean. Crop Science, 31, 88-89.

Flores, T., Karpova, O., Su, X., Zeng, P., Bilyeu, K., Sleper, D. A., et al. (2008). Silencing of *GmFAD3* gene by siRNA leads to low α-linolenic acids (18: 3) of *fad3*-mutant phenotype in soybean [*Glycine max* (Merr.)]. Transgenic Research, 17, 839-850.

Gillman, J. D., Tetlow, A., Hagely, K., Boersma, J. G., Cardinal, A., Rajcan, I., et al. (2014). Identification of the molecular genetic basis of the low palmitic acid seed oil trait in soybean mutant line RG3 and association analysis of molecular markers with elevated seed stearic acid and reduced seed palmitic acid. Molecular Breeding, 34, 447-455.

Haun, W., Coffman, A., Clasen, B. M., Demorest, Z. L., Lowy, A., Ray, E., et al. (2014). Improved soybean oil quality by targeted mutagenesis of the fatty acid desaturase 2 gene family. Plant Biotechnology Journal, 12, 934-940.

Hyten, D. L., Pantalone, V. R., Saxton, A. M., Schmidt, M. E., and Sams, C. E. (2004). Molecular mapping and identification of soybean fatty acid modifier quantitative trait loci. Journal of the American Oil Chemists' Society, 81, 1115-1118.

Kinney, A. J. (1995). Improving soybean seed quality. in Induced mutations and molecular techniques for crop improvement. (International Atomic Energy Agency), 101-113.

Kulkarni, K. P., Kim, M., Song, J. T., Bilyeu, K. D., and Lee, J. D. (2017). Genetic improvement of the fatty acid biosynthesis system to alter the ω-6/ω-3 ratio in the soybean seed. Journal of the American Oil Chemists' Society, 94, 1403-1410.

Lee, J. D., Bilyeu, K. D., and Shannon, J. G. (2007). Genetics and breeding for modified fatty acid profile in soybean seed oil. J Crop Sci Biotech, 10, 201-210.

Li, Y., Hruby, A., Bernstein, A. M., Ley, S. H., Wang, D. D., Chiuve, S. E., et al. (2015). Saturated fats compared with unsaturated fats and sources of carbohydrates in relation to risk of coronary heart disease: a prospective cohort study. Journal of the American College of Cardiology, 66, 1538-1548.

Nyquist, W. E., and Baker, R. J. (1991). Estimation of heritability and prediction of selection response in plant populations. Critical Reviews in Plant Sciences, 10, 235-322.

Okuley, J., Lightner, J., Feldmann, K., Yadav, N., and Lark, E. (1994). Arabidopsis *FAD2* gene encodes the enzyme that is essential for polyunsaturated lipid synthesis. The Plant Cell, 6, 147-158.

Pham, A. T., Lee, J. D., Shannon, J. G., and Bilyeu, K. D. (2010). Mutant alleles of *FAD2-1A* and *FAD2-1B* combine to produce soybeans with the high oleic acid seed oil trait. BMC Plant Biology, 10, 195.

Pham, A. T., Shannon, J. G., and Bilyeu, K. D. (2012). Combinations of mutant *FAD2* and *FAD3* genes to produce high oleic acid and low linolenic acid soybean oil. Theoretical and Applied Genetics, 125, 503-515.

Pham, A. T., Bilyeu, K., Chen, P., Boerma, H. R., and Li, Z. (2014). Characterization of the *fan1* locus in soybean line A5 and development of molecular assays for high-throughput genotyping of *FAD3* genes. Molecular Breeding, 33, 895-907.

Primomo, V. S., Falk, D. E., Ablett, G. R., Tanner, J. W., and Rajcan, I. (2002). Inheritance and interaction of low palmitic and low linolenic soybean. Crop Science, 42, 31-36.

Ross, A. J., Fehr, W. R., Welke, G. A., and Cianzio, S. R. (2000). Agronomic and seed traits of 1%-linolenate soybean genotypes. Crop Science, 40, 383-386.

Sandhu, D., Alt, J. L., Scherder, C. W., Fehr, W. R., and Bhattacharyya, M. K. (2007). Enhanced oleic acid content in the soybean mutant M23 is associated with the deletion in the *Fad2-1a* gene encoding a fatty acid desaturase. Journal of the American Oil Chemists' Society, 84, 229-235.

Shah, S., and Xin, Z. (1997). Overexpression of the *FAD3* desaturase gene in a mutant of Arabidopsis. Plant Physiology, 114, 1533-1539.

Sweeney, D. W., Carrero-Colón, M., and Hudson, K. A. (2017). Characterization of new allelic combinations for high-oleic soybeans. Crop Science, 57, 611-616.

Thapa, R., Carrero-Colon, M., Crowe, M., Gaskin, E., and Hudson, K. (2016). Novel *FAD2–1A* alleles confer an elevated oleic acid phenotype in soybean seeds. Crop Science, 56, 226-231.