6-1-2021

Identifying Frogeye Leaf Spot Resistance in Two Elite Soybean Populations and Analysis of Agronomic Traits in Resistant Lines

Kelsey Smith
Southern Illinois University Carbondale, 2015ksmith@gmail.com

IDENTIFYING FROGEYE LEAF SPOT RESISTANCE IN TWO ELITE SOYBEAN POPULATIONS AND ANALYSIS OF AGRONOMIC TRAITS IN RESISTANT LINES

by

Kelsey Smith

B.S., Southern Illinois University, 2019

A Thesis
Submitted in Partial Fulfillment of the Requirements for the
Master of Science Degree

Department of Plant, Soil, and Agricultural Systems
in the Graduate School
Southern Illinois University Carbondale
May 2021

THESIS APPROVAL

IDENTIFYING FROGEYE LEAF SPOT RESISTANCE IN TWO ELITE SOYBEAN POPULATIONS AND ANALYSIS OF AGRONOMIC TRAITS IN RESISTANT LINES

by

Kelsey Smith

A Thesis Submitted in Partial

Fulfillment of the Requirements

for the Degree of

Master of Science

in the field of Plant, Soil, and Agricultural Systems

Approved by:

Dr. Stella Kantartzi, Chair

Dr. Ahmad Fakhoury

Dr. Andrew Wood

Graduate School
Southern Illinois University Carbondale
March 24, 2021

AN ABSTRACT OF THE THESIS OF

Kelsey Smith, for the Master of Science degree in Plant, Soil, and Agricultural Systems, presented on March 24, 2021, at Southern Illinois University Carbondale.

TITLE: IDENTIFYING FROGEYE LEAF SPOT RESISTANCE IN TWO ELITE SOYBEAN POPULATIONS AND ANALYSIS OF AGRONOMIC TRAITS IN RESISTANT LINES

MAJOR PROFESSOR: Dr. Stella Kantartzi

Soybeans (*Glycine max* L.) are an important crop globally for its food, feed, and oil purposes. It is impacted by many diseases, including *Cercospora sojina*, the causal agent of Frogeye Leaf Spot (FLS). Chemical and cultural controls to this fungal pathogen are insufficient, so genetic resistance must be acquired for adequate control. To this end, two recombinant inbred populations were screened in a greenhouse setting for their relative resistance to FLS, and their genomes were analyzed for contributing quantitative trait loci (QTL). In the Essex × Forrest population, one QTL was discovered on chromosome 13, and in the Forrest × Williams 82 population, two QTL were identified on chromosomes 6 and 11, respectively. These populations were then also screened in a field setting for agronomic traits. These traits were analyzed to detect one superior line for both FLS resistance and advanced agronomic traits, F×W 125. This line should be used in future breeding projects to increase FLS resistance and reduce linkage drag for other desired characteristics.

ACKNOWLEDGMENTS

A sincere thank you goes to Dr. Stella Kantartzi for the mentorship she has provided over the last two years. Without your scholarship and direction, none of the work outlined in this thesis would have ever been possible. I would also like to acknowledge Yi-Chen Lee for the help he provided in greenhouse experiments, field trials, and statistical analyses. A great deal of gratitude goes to Dr. Ahmad Fakhoury and Dr. Andrew Wood for serving on my thesis committee and dedicating their time and talents to my thesis defense. Finally, a sincere thank you goes to the entire School of Agricultural Sciences at Southern Illinois University Carbondale. Many faculty have contributed to my academic success over my six years at SIU, and I am truly grateful for the well-rounded education I received in agricultural science.

DEDICATION

There are so many people in my personal life that have supported me in this journey. To start, I am thankful for my husband, Austin, who has been incredibly supportive of my educational goals. Thank you for listening to every moment of inspiration and frustration I have had while digging into the world of plant genetics, even when I sounded like I was speaking a different language. I sincerely thank my parents, for raising me to believe I could do absolutely anything I set my mind to and instilling in me a hard work ethic. To my brother, Collin, thank you for being willing to have two-hour phone calls about work, life, and our careers. You are the only other person who shares a gene pool with me, and you have always pushed me to reach higher. To my best friend, Josie, thank you for going through this crazy ride called graduate school with me. The last six years at SIU would have never been as amazing without you as a roommate and fellow grad student. Finally, to my friends Anastasia, Tristan, Mindy, and Christian, thank you for dealing with all of my antics and providing support, humor, and friendship for all of these years. It truly takes a village, along with a lot of coffee.

TABLE OF CONTENTS

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

FIGURE PAGE

CHAPTER 1

REVIEW OF LITERATURE

A. Soybean History and Physiology

Soybeans *(Glycine max* L. Merrill) are a legume plant used globally for its edible seed. Also known as soja bean or soya bean, this crop is one of the cheapest sources of protein in the world and is a staple in millions of people and animals' diets globally (Britannica, 2019). The origins of soybeans are mottled, but it is believed to have originated in northeast China. These soybeans were smaller than current beans and had a black seed coat, much like the 'Peking' variety has today (Singh 2010). It is believed that soybeans were first domesticated in China as early as 7000 BC. From there, it moved to Japan and was first cultivated around 5000-3000 BC (Lee et al., 2011). Edamame, soy sauce, tofu, and other soybean-based products are still quite popular in these regions due to the crop's abundance (Britannica, 2019). From East Asia, soybeans migrated to Indonesia, the Philippines, Vietnam, Thailand, Malaysia, Burma, Nepal, and North India, which are known as the secondary gene centers of soybeans. This movement was in large part due to sea and land trade routes that were being newly created, such as the Silk Road (Hymowitz, 1990). In 1765, the surveyor general of Georgia named Henry Yonge planted what he referred to as 'Chinese vetch' after receiving the seeds from a merchant at the East India Trading Company (Yonge, 1767). This would be the first patch of soybeans planted in the United States, and the crop was widely distributed to the Midwest by 1882 (Singh 2010). Over time, the crop has been introduced to much of the world, and it is currently grown wherever conditions are adequate.

Soybeans are successfully grown in regions where summers stay between 20-30°C (Singh 2010). They grow well in soils with a pH of 6.0-7.0, with 6.3-6.5 being the most ideal

range for nutrient availability (Staton, 2012). This makes it an ideal crop for the Northern Hemisphere, but difficult for countries in the Southern Hemisphere. The added heat and humidity of tropical regions used to hinder growth and production. The plants would also reach flowering during the summer solstice, when determinate varieties reach maximum height. This caused early pod set and reduced yields. After many years of crossbreeding, tropical soybean varieties were developed, allowing Brazil to become a top-producing country of soybeans (Alves, Boddey, & Urquiaga, 2003).

In 2018, Brazil was the top soybean-producing country in the world with 126 million tons, with the United States closely following with 124 million tons. There is a sharp drop off after the US, with Argentina, China, and India producing 38, 14, and 14 million tons, respectively (Food and Agriculture Organization of the United Nations, 1997). Eighty-five percent of all soybeans produced will go to animal feed, with the remaining 15% being used for many edible and non-edible products such as soy-based foods, biodiesel, cooking oil, and industrial applications. Globally, 80% of all soybean production comes from large-scale farming operations, and the remaining 20% comes from small-scale farmers (Voora, Larrea, & Bermudez, 2020). Demand continues to grow yearly as biodiesel demands increase, along with global meat consumption. In Western countries where meat consumption is decreasing, it is expected that soy-based alternatives will take their place in the market, making it a stable commodity for farmers to grow regardless of market status (Voora, Larrea, & Bermudez, 2020).

Soybean seeds are pulses that are comprised of epicotyl that will form the shoot and leaves, a radicle that becomes the roots, a hypocotyl that connects the cotyledon and radicle, cotyledons that act as a food source and beginning leaves, and a seed coat to protect the seed (Singh, 2010). When the seed is exposed to optimal moisture and temperature, germination

begins. The seed will double in size and increase its seed moisture by 50%. Respiration increases, and in ideal temperatures, the radicle emerges in four days. Next, the hypocotyl emerges and the two cotyledons fold out and act as temporary leaves for the plant. Over time, these cotyledons will fall off and the true leaves will take over photosynthesis (Purcell, Salmeron, & Ashlock, 2014). Next, the soybean will produce one set of unifoliate leaves, and all following sets of leaves are grown as trifoliates. When the first trifoliate emerges, the plant is in the V1 (Vegetative 1) stage. The V2 stage is when the second trifoliate emerges, and this system of growth continues until flowering.

Once blooms begin to form on any node of the main stem, the plant is said to be in Reproductive 1 (R1) phase. Reproductive phases also continue numerically, with R2 being full bloom, R3 being beginning pod, R4 being beginning seed, R6 being full seed, R7 being beginning maturity, and R8 being full maturity (Purcell, Salmeron, & Ashlock, 2014). Soybeans are self-pollinating with either white or purple flowers, depending on the variety. Flowering and pod set are short-day photoperiod sensitive and begin flowering when days are shorter than twelve hours (Destro, Carpentieri-Pipolo, Kihl, & Almeida, 2001). Two types of growth patterns exist in soybeans as well: determinate and indeterminate. Determinate plants halt lateral growth once flowering begins, and indeterminate plants continue to grow post-flowering (Bernard, 1972). In this way, yield, growth pattern, and time to flowering are directly intertwined.

Another particularly intriguing characteristic of soybeans is their ability to fix nitrogen from the soil. Nitrogen is abundant in the soil in a form unusable to most crops, and therefore fertilizers must be applied. Soybeans, like many other legumes, have nodules on the roots that can convert soil nitrogen into a usable form. This happens through a symbiotic relationship between the soybean plant and *Bradyrhiyzobium japonicum*, a type of rhizobacteria (Miransari,

2016). Much of the time, soybeans do not need any assistance in this task. However, studies have shown that when *B. japonicum* is added to the environment, soybeans can be even more productive at nitrogen fixing (Elhady, Hallman, & Huer, 2020). After soybeans have been removed from a field, nitrogen in its usable form is left behind, making soybean an ideal crop to be grown in rotation with corn. Corn is infamous for using drastic amounts of nitrogen during growth, and applications of nitrogen can become costly. When planted in a field after soybeans, this cost can be reduced significantly (Laur, Porter, & Oplinger, 1997).

B. Diseases of Soybeans

Pathogens effecting soybeans come in various forms. Fungi, bacteria, viruses, and nematodes all use soybeans as their host, causing damage to the plant and farmers' yield. Hundreds of thousands of bushels of soybeans are lost to soybean diseases yearly, though farmers tend to underestimate how much they are losing to disease (Allen et al., 2017). Instead, producers tend to view weeds as the top stealer of yield, likely because diseases can commonly be an invisible enemy (Aref & Pike, 1998).

Fungal organisms can attack soybean plants in many ways. They can attack as soilborne pathogens, root diseases, stem diseases, or foliar diseases (Boerma & Specht, 2004). There are also various classifications of fungal diseases based on spore type: ascomycetes, basidiomycetes, deuteromycetes, oomycetes, and zygomycetes being the most common (Cooper, 2007). Fungal diseases are the most common type of diseases in plants, with sudden death syndrome, frogeye leaf spot, and charcoal rot being the most destructive on soybean in 2014 (Allen et al., 2017). Fungal life cycles are complex, and they can reproduce sexually or asexually depending on the type of disease and where it is in its life cycle. Plant-pathogenic fungi usually enter through an open wound in the plant, or they bore through the cuticle of the plant by utilizing a specialized

structure called an appressorium at the end of a germ tube. As turgor pressure builds along the appressoria, an infection peg is used to breach plant cells and infect the plant (Carrls, Little, & Stiles, 2012). Cultural control of fungal pathogens is largely dependent on what type of disease it is and its life cycle. Chemical control exists for many economically important species, as well as genetically modified soybean varieties as a method of host resistance (Boerma & Specht, 2004).

Bacteria are less common sources of disease, but nonetheless important. Bacterial blight (*Pseudomonas syringae* pv. *glycinea)*is the most common bacterial disease, followed by bacterial pustule and bacterial wilt (*Ralsonia solanacearum*) (Boerma & Specht, 2004). Bacterial diseases do not normally devastate crops and fields, though bacterial wilt has done considerable damage in the Ukraine (Hartman et al., 1999). Like fungi, most bacteria enter through a wound on the plant. Bacterial cells can also travel on wind-driven rain, allowing the pathogen to spread for miles. Bacterial diseases can overwinter on seed, weed residue, and crop residue. For this reason, cultural control generally includes removal of prior crop residues, weed management, and the use of clean seed (Boerma & Specht, 2004). Crop rotation can also be used as a cultural method by planting non-hosts between hosts so that cells cannot survive in the field for a year. Very few genetically resistant soybean lines have been implemented for bacterial diseases (Boerma & Specht, 2004).

Viral pathogens are generally carried on vectors such as aphids. Viruses most typically show early symptoms on the leaves as a mosaic, mottle, or chlorosis. However, they can also show symptoms on stems, pods, and seeds (Boerma & Specht, 2004). There are fifteen viruses that commonly effect soybeans in the US, with the most common genera being *Potyvirus*, *Comovirus*, and *Nepovirus*. Identifying viral infections can be quite challenging. Sometimes they can be identified based on symptomology, but many times serological methods like enzyme-

linked immunosorbent assays (ELISA) tests must be conducted. Other times polymerase chain reaction (PCR) tests or genotyping must be done to conclusively determine which virus has infected a field (Boerma & Specht, 2014). So far, viruses have not caused severe economic impact in the United States, so cultural methods to reduce insect vectors have been historically sufficient. However, there is continuing research being conducted to find genes of resistance to implement host resistance (Hill & Whitham, 2014).

Nematodes are one of the most economically important types of soybean pest, following fungi. The major pathogenic nematode in the world is soybean cyst nematode (*Heterodera glycines*). Others include lance nematodes, root-lesion nematodes, and reniform nematode, but their economic importance is paled in comparison to *H. glycines* (Boerma & Specht, 2014). Cyst nematode was the most damaging plant pathogen, and cost farmers of the most yield in the 2010-2014 growing seasons (Allen et al., 2017). Adult female cyst nematodes burrow into soybean roots where she is fertilized by a male and remains until her death. It is at this point that she is a cyst, which can be seen on the root surface with the naked eye. As juveniles hatch, they break through the body of the deceased mother, and the cycle continues. The entire life cycle can happen in as little as four weeks (Atibalentja, Jakstys, & Noel, 2004). The major cultural method available for cyst nematode has been non-host rotation, but this alone is not effective enough due to this species' ability to lie dormant for years on end (Niblack, 2007). Genetic resistance is implemented widely with the utilization of the *rhg1-b* allele, and work continues to stack resistance genes (Cook et al., 2012).

C. Frogeye Leaf Spot

Frogeye leaf spot of soybeans (FLS) is caused by a fungal pathogen known as *Cercospora sojina.* It is signified by lesions that start out small, gray, and water-soaked. As the

fungus spreads, these lesions can coalesce and gain a dark red-brown border. When infection has been present for some time, conidiophores can be detected with the bare eye in the center of the lesion (Lin & Kelly, 2018). While FLS is generally a foliar disease, it can spread to stems, pods, and seeds in the late growing season. This loss of photosynthetic tissue reduces the productivity of the plant, and can cause the plant to wither and fall prematurely in severe infections (Lin & Kelly, 2018).

The first case of FLS was reported in the United States in 1924 (Boerma & Specht, 2004). Historically, this disease was most prevalent in the southern United states, but it has spread north in recent years, with reported cases as far north as Wisconsin (Mengistu, Kurtzweil, & Grau, 2007). In the southern United States, Frogeye Leaf spot was listed as one of the top five most destructive soybean diseases during the 2012-2014 growing seasons (Allen et al., 2017). In highly infected fields, yield losses have been estimated as high as 60% (Mengistu, Kurtzewil, & Grau, 2007). As northern states have warmer weather, it is expected that northern states will have higher incidence and yield loss from this disease.

Historically, there were five well-understood races of *C. sojina,* races 1-5. It was well understood that there were likely many more races that had not been characterized, with an acceptance that there were likely at least twelve races in the U.S. Brazil reported 22 races, and China reported 14 races when trying to characterize genetic variability (Boerma & Specht, 2004). In 2007, eleven new race designations were proposed. These races are known as race 5-15, and are the currently accepted race designations (Mian, Missaoui, & Walker, 2007).

C. sojina has a life cycle that is repeated throughout a growing season. In this way, the more fungal spores present in the primary inoculum greatly determines how severe an infection will be. *C. sojina* thrives in high humidity (>90%) and the spores spread through rain and dew.

After infection, lesions can show on leaves in 7-14 days (Lin & Kelly, 2018). Conidiophores grow on the leaf tissue, and conidia are dispersed to repeat the life cycle. If it is the end of the growing season, the fungus can instead overwinter on leaf tissue and other plant parts. Next growing season, the conidia that overwintered will become the primary inoculum in the next year (Lin & Kelly, 2018).

Cultural control methods include crop rotation to prevent overwintered spores from having a host. Tillage shows minor benefits, but does not greatly reduce inoculum from year to year. Planting inoculum-free seed is also integral to preventing large outbreaks, though it is believed plant residue is the most important source of inoculum (Lin & Kelly, 2018). Quinone outside inhibitor fungicides (FRAC group 11) were historically used to control *C. sojina,* but resistance has been obtained by the pathogen. Resistance had been detected in the southern United States as early as 2010, and had reached as far north as Iowa in 2017 (Zhang et al., 2018).

There are three main genes of resistance that confer resistance: *Rcs1, Rcs2,* and *Rcs3.* These resistance genes distinguish which races of *C. sojina* they confer resistance to in the original races 1-5. *Rcs1* confers resistance to race 1, *Rcs2* confers resistance to race 2, and *Rcs3* confers resistance to all other known races (Mian et al., 2007). Work is still currently being done to distinguish how these resistance genes transfer to the new race 5-11 system. In 2012, two resistance alleles were identified: *Rcs* (PI 594891) and *Rcs* (594774). It is believed that these two alleles are related to the *Rcs3* gene, but more work is being done to understand the exact relationship (Pham et al., 2015).

D. Plant Breeding

The science of plant breeding is defined as the improvement of plant genetics through crossing plants that have desired traits and selecting progeny plants with improved combinations

of traits (Chahal & Gosal, 2002). It is a science of targeted evolution that also requires an understanding of how plants interact with their environment. The genotype for a trait interacts with its environment to create a phenotype that is either desired or undesirable (Chahal & Gosal, 2002). Much of plant breeding today focuses on adapting plants to biotic or abiotic stresses. Biotic stresses include diseases, insects, and weed pressure. Abiotic stressors can be drought, flooding, poor soil conditions, and other climate or nutrient related stressors. Creating plants that can tolerate all of these stresses allow for farmers to produce crops in a stable manner every year, regardless of what challenges the growing season provides (Chahal & Gosal, 2002).

Traditional methods of plant breeding include sexual hybridization, wide crossing, tissue culture, and mutagenesis (Chahal & Gosal, 2002). Sexual hybridization is the crossing of two plants either in nature, or selectively by breeders to create progeny. It is generally imprecise, and is the most traditional breeding method (Negrutiu et al., 1989). The most popular hybridization methods are single crossing with two parents, three way crosses with three parents over two generations, four way crosses with four parents over two generations, and complex crosses with greater than four parents (Cahal & Gosal, 2002). Wide crossing uses genetic material from outside of that particular species to add genetic variation. This can be done with historic versions of modern crops, and alters the genome in ways traditional crossing cannot (Yang et al., 2020). Tissue culture is a process of maintaining plant cells and tissues in a lab for the purpose of creating new plants from functional cells (Cahal & Gosal, 2002). Mutagenesis is the process of using chemicals known as ‘mutagenic compounds’ that can alter the genome in a targeted way either through radiation or chemical reactions to create an intended change (Ling & Robinson, 1997). From here, seeds can be grown in the greenhouse or field to examine which plants have desired characteristics.

A more modern approach to plant breeding is the use of marker-assisted selection (MAS). Molecular markers are a gene with known functions and locations that can be used to study the genes around it and the inheritance of that gene. These markers can be either DNA markers, or protein markers (Cahal & Gosal, 2002). There are different types of markers, with varying levels of precision. From least to most precise, the most common markers used are allozymes, random amplified polymorphic DNA (RAPD), amplified fragment length polymorphisms (AFLPs), multi-locus fingerprints (RFLPs), microsatellites (SSRs), and single nucleotide polymorphisms (SNPs) (Cahal & Gosal, 2002). Allozymes are a protein electrophoresis-based marker, and are the least precise method. These proteins are detected using electrophoresis, and signify differences among alleles (Krause & Brand, 2016). RAPDs are a marker that was developed by amplifying random sequences of markers through random primers. It is quick and easy, but it is not easily reproduced (Cahal & Gosal, 2002). AFLPs use the cleavage of two DNA fragments with enzymes to amplify subsets of the DNA. This polymerase chain reaction (PCR) product is then is separated on an acrylamide gel and combinations of links are readily available to analyze the results. It is by far the most complicated method with only moderate sensitivity (Cahal & Gosal, 2002). RFLPs use a restriction enzyme to recognize variation in a DNA sequence. Gel electrophoresis is used to visualize the RFLPs, and the bands shown on the gel indicate nucleic acids present (National Human Genome Research Institute, 2020). SSR satellites are used by amplifying specific known sequences in the genome to view it on a gel configuration. It is very precise, but can be expensive (Cahal & Gosal, 2002). Finally, SNPs use differences of a single base pair as markers. Any two individuals that are unrelated usually have one base pair difference every 1,000 base pairs, with no effect on cell function. This makes it an incredibly sensitive marker (Cahal & Gosal, 2002). SNP markers are now considered to be the gold

standard of MAS for its sensitivity, relative cheapness, and ease of use (Mammadov, Aggarwal, Buyyarapu, & Kumpatla 2012).

Quantitative trait loci (QTL) are defined as a portion of a chromosome that can be identified by a molecular marker and has a significant effect on a quantitative trait. The use of QTL related to a desired trait is much cheaper, faster, and more efficient than large phenotypic assays. Any marker type can be related to a QTL, and using lab techniques to identify what traits a seed holds before planting it can speed up the breeding timeline of a given project. The use of QTL removes many random errors and can be coupled with field and greenhouse assays to quickly produce high-performing lines quickly and efficiently (Yin, Stam, Kropff, & Schapendonk, 2003).

Recombinant inbred lines (RILs) are another powerful tool plant breeders have in modern genetic mapping. RIL populations are created by crossing two parents to create an F_1 generation with half of the genome from each parent. These progeny are then self-pollinated for five or more generations to reach a stable genome (Broman 2005). This creates a population with mixed genetics of both parents. A normal distribution should occur at any given QTL, and extremes on either end of the distribution can lead researchers to new discoveries. Coupled with QTL and SNP methods, RILs are an effective way for plant breeders to reach their research goals (Broman 2005).

CHAPTER 2

PINPOINTING FLS RESISTANCE IN TWO SOYBEAN POPULATIONS BY UTILIZING QTL

'Essex' × 'Forrest' data has been published in "QTL mapping for resistance to *Cercospora sojina* in 'Essex' × 'Forrest' soybean (*Glycine max* L.) lines". *Journal of Plant Breeding and Crop Science 13*(1): 14-22.

ABSTRACT

Frogeye leaf spot (FLS), caused by *Cercospora sojina* Hara, is observed as red-brown lesions on leaves that can coalesce and decrease the photosynthetic ability of soybeans. The average yield loss due to Frogeye Leaf Spot is estimated at approximately 40% in established fields, whereas 100% incidence was previously recorded. Quinone outside inhibitor fungicides were considered an effective control method, but the pathogen quickly evolved an ability to thrive post-application. This trait quickly spread across North America. Therefore, genetic host resistance is likely the most effective method to prevent the disease. To achieve this goal, we aimed to screen 91 recombinant inbred lines (RILs) of 'Essex' × 'Forrest' and 190 RILS of 'Forrest' × 'Williams' under greenhouse conditions for FLS resistance and used single nucleotide polymorphism (SNP) markers to identify associated quantitative trait loci (QTL). Three QTL were mapped in this study. In 'Essex' × 'Forrest', one QTL was reported on Chr. 13, and in 'Forrest' × 'Williams 82' two QTL were reported on Chr. 6 and Chr. 11. Overall, this study will help to better understand the underlying mechanisms of soybean resistance to *C. sojina* Hara as well as to develop soybean varieties with resistance to FLS using marker assisted selection.

Keywords: Cercospora sojina Hara, quantitative trait loci, Frogeye Leaf Spot, Essex × Forrest, Forrest × Williams 82, disease resistance, genotypic and phenotypic traits

INTRODUCTION

Frogeye leaf spot (FLS), caused by the pathogen *Cercospora sojina* Hara, is a foliar disease indicated by water-soaked lesions on the leaves of soybeans. The lesions begin as small brown spots and develop a dark, red-brown border, whereas in severe cases, they can also form on the stems, pods, and seeds. When lesions appear on seeds, the fungus spreads to new seedlings the following year (Malvick, 2018). Yearly soybean losses to FLS in the United States have been measured at 106.3 thousand metric tons, with the most losses in the southern states (Wrather et al., 2001). In heavily infected fields, FLS can reduce soybean yield by 40% in conducive environmental conditions (Byamukama et al., 2019). Together, these characteristics create a cycle of reduced yield and reduced profits for infected fields.

The first verified case of FLS in the United States of America was recorded in 1925 (Lehman, 1928). The disease was particularly problematic in the southern states for many years, with cases first recorded in the Midwest in the late 1940s (Philips and Boerma, 1981). For many years, chemical control, mostly using uinone outside inhibitor (QoI) fungicides (also known as FRAC Group 11), was the most effective method for disease management. FLS resistance to QoI inhibitors was detected in North America by 2010 (Zhang, 2012), making genetic host resistance to FLS more crucial to high-yielding soybean production.

Single nucleotide polymorphisms (SNPs) for disease resistance in soybean are usually centralized on chromosomes (Chr.) 7, 13, and 18. Chr. 13, in particular, is known to be a rich area of disease resistance, as it harbors the resistance gene rich Satt114 marker and the *Rsp8* gene. This area is associated with resistance to two races of *Phytophthora sojae*, the causal agent of Phytophthora root rot. (Gordon et al., 2006). Satt114 is also commonly used as a flag marker for other disease resistance studies (Pham et al., 2015). However, resistance genes are not

restricted to these areas and can be scattered across the genome. For example, SNPs that are significant to Soybean cyst nematode resistance can be found on Chr 3, 4, 7, 9, 10, 11, 13, 14, 15, 18, 19, and 20 (Chang et al., 2016).

Currently, there are 12 known races of *C. sojina* Hara and three main genes conferring resistance. These genes are *Rcs1*, which codes for resistance to race 1; *Rcs2*, which provides resistance to race 2; and *Rcs3*, which confers resistance to all other known races of *C. sojina* Hara (Mian et al., 2007). In 2012, two additional dominant resistance alleles were identified as *Rcs* (PI 594891) and *Rcs* (PI 594774) (Pham et al., 2015). More research is needed in this area to understand specific QTL that are associated with each resistance gene to make their implementation more feasible for breeders. New race designations were also implemented in 2007, with the new races being races 5-11. Work is still being done to associate the known resistance genes with the new race designations (Mian et al., 2007).

The 'Essex' × 'Forrest' (E × F) cross was made at Southern Illinois University Carbondale to be a mapping population for a variety of traits (Lightfoot et al., 2005). Essex was chosen for its partial resistance to FLS, whereas Forrest for its partial susceptibility (Sharma and Lightfoot, 2017). Forrest has been extensively studied and mapped alongside 'Williams 82', making it an ideal candidate line for QTL identification. Essex and Forrest share a common germplasm heritage that accounts for 25% of their genomes. (Lightfoot, 2008). From the initial cross, approximately 4,500 F_2 plants were advanced to F_5 using single-pod descent. After harvest, 150 F_5 plants were randomly selected and planted into progeny rows. Of these, 100 recombinant inbred lines (RILs) were kept for various phenotypic assays. In total, 94 RILs were used to construct a mapping population for quantitative trait loci (QTL) discovery and also

released for research purposes (Lightfoot et al., 2005). The plant material used in this study consisted of 91 $F_{5:8}$ selected RILs.

The 'Forrest × Williams 82' population was created by crossing 'Forrest' × 'Williams 82' to create F1 seeds. This generation was advanced to F_2, and each F_2 plant was advanced to F_7 by the single seed descent (SSD) method. In the F_8 generation, the $F_{2:7}$ seeds were bulked in 1m rows to create 1,025 $F_{2:7}$ recombinant inbred lines (RILs) to be used for genetic mapping (Wu et al., 2011). Of these lines, 190 were used and maintained at Southern Illinois University for this study.

Markers closely linked to QTL can be used to screen hundreds of lines at once for the genes of interest. For the purpose of developing resistant cultivars, the use of marker assisted selection is an efficient and accurate way to identify resistant lines as opposed to large phenotypic surveys (Yousef and Juvik, 2001). Phenotypic assays require more labor, take longer to complete, and are less precise compared to genotypic methods. Two major QTL for FLS resistance were detected in the E × F population for *C. sojina* Hara race 2 on Chr. 6 near Satt319 and on Chr. 8 near Satt632 as well as 13 minor QTL across various chromosomes (Sharma and Lightfoot, 2017). However, this study used simple sequence repeat (SSR) to find regions of interest. The use of SNP markers are more precise than SSR and are the preferred method in genetic diversity studies (Singh et al., 2013). For this reason, SNP were used in this study. Having a precise location in the genome for FLS resistance allows for simpler implementation in commercial lines.

The objectives of this study were to analyze the phenotypic variation of FLS resistance in E×F in a greenhouse setting, create a genetic linkage map for the population, and identify candidate QTL that code for resistance to *C. sojina* Hara race 15 using SNPs.

MATERIALS AND METHODS

A. Greenhouse Assay

Greenhouse assays were conducted by planting one population and their parental lines in six-inch plastic nursery pots filled with Berger BM1 growing medium. Soybean plants were grown in open benches with no supplemental lighting and were allowed to experience ambient conditions. Plants were watered according to environmental needs, generally twice a week, with tap water. No fertilization was used in this experiment. Pots were arranged in a randomized complete block design with two blocks per replication. Each population was replicated twice in time, with the E × F experiments taking place in March 2019 and October 2019, and F × W experiments taking place in August 2019 and February 2020. Temperatures stayed between 18-35°C over the duration of the experiments, Seven seeds were planted in each pot. One treatment, the application of *C. sojina* Hara spores, was applied to all blocks. Shortly after emergence, thinning was performed to a density of one plant per pot. Plants were inoculated for the first time with *C. sojina* Hara solution at V2–V5 stages. Plants were then inoculated a second and third time with a week between inoculations.

Race 15 of *C. sojina* Hara was cultured in petri dishes filled with clarified V8 solid medium (Salas et al., 2007). After two weeks in a growth chamber at 25°C, the petri dishes were flooded with a 0.1% Tween 20 solution and spores were knocked into the solution using a sterilized metal spatula. Approximately eight petri dishes of seven colonies were used to make 300 mL of solution. The solution was mixed thoroughly on a stirring plate for 5 min, and then was filtered through a cheesecloth to remove mycelium. Final spore concentration was approximately 6 x 10^4 conidia/mL. This final product was poured into a spray bottle and immediately used for inoculation.

All lines were sprayed to dripping with the fungal solution and covered using a gallon-sized plastic bag to create a highly humid microenvironment. Gallon-sized bags were left on for 72 h. For the rest of the experiment, the plants were left under a humidity tent using plastic sheeting and a humidifier. Relative humidity was maintained at 80–90% and temperature was maintained at 28–30°C until the end of the experimental period.

Two weeks after the first inoculation, plants were rated for disease severity using a numeric scale from 1-10. This method allowed for characterization of disease development over time. Plants were rated on a scale of 1–10; rating of 1 indicates 0–10% of the leaf surface showing disease symptoms, whereas a rating of 10 indicates 90–100% of the leaf showing symptoms. Defoliation due to disease presence was also counted as a 10 (Sinclair, 1982). In total, six ratings were taken within 2 wks.

B. DNA Isolation

For DNA isolation, all lines screened in the greenhouse were planted in six-pack trays and allowed to grow in a dark room to minimize cuticle growth and chloroplastic DNA expression. When plants reached the V1 stage (first trifoliate emergence), 50 mg of tissue from the first trifoliate was collected and stored in a -20°C freezer until isolation. Upon collection of all tissues, samples were thawed, flash frozen with liquid nitrogen, and crushed. DNA isolation was performed using the DNeasy 96 Plant Kit (Qiagen, Hilden, Germany), following the manufacturer's instructions. DNA purity was tested using a gel electrophoresis visualized with a 1% EtBr stained agarose gel, and DNA quantification was carried out with NanoDrop 2000 (Thermo Scientific, Waltham, MA, USA). SNP genotyping was conducted at the Soybean Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD, using the BARCSoySNP6K BeadChip array.

C. Phenotypic Variation

To compare FLS resistance across the population, the sixth and final greenhouse rating for each line was used to run a distribution analysis. Lines with a lower FLS score than the susceptible parent were labelled "susceptible lines" and lines with higher FLS scores than the resistant parent were labelled "resistant lines." JMP Pro 15 software was utilized to run distributions and average disease scores.

D. Genetic Map and QTL Analysis

The genetic map and QTL analysis were done with the r/QTL package for R Studio(Broman et al., 2003; Broman and Sen, 2009). The final rating for each line was used to measure the overall FLS resistance. Frogeye leaf spot scores were used to find phenotypic and genotypic differences between the parental lines and the RILs. Single marker analysis and interval mapping were used to identify the chromosomes of interest (data not shown), the Cim() function was subsequently used for composite interval mapping (CIM). The Fitqtl() function was used to estimate the variance of QTL of interest, and a 1,000 permutation test was ran to determine approximate logarithm of odds (LOD) thresholds of significance using *operm.ag*.

E. Gene Ontology and Kyto Encyclopedia of Genes and Genomes Pathways

The SoyBase database (Wm.82 version 1) was utilized to analyze the gene ontology (GO) and kyto encyclopedia of genes and genomes (KEGG) pathway of the candidate QTL and identify which proteins are coded for in the CIM interval. The UniProt Consortium database was then used to understand what these proteins do within the plant so that overall gene function can be understood.

RESULTS

A. Phenotypic Variation

The distribution of FLS scores across E × F population was normal (P=0.158), the kurtosis of the distribution was 0.004 and the skewness was 0.31. Overall, the average of the FLS score was 3.23 ± 1.32, and the scores ranged from 1 to 7.25. Five lines were identified as more resistant than Essex (average score, 1.50 ± 0.50), whereas two lines were more susceptible than Forrest (average score, 5.75 ± 2.49) (Fig 1.). Lines more resistant than Essex were noted as E × F 2, E × F 9, E × F 10, E × F 11, and E × F 54 (average score, 1.0 ± 0). The lines more susceptible than Forrest were E x F 29 (average score, 7.25 ± 1.79) and E × F 63 (average score, 6.0 ± 2.0).

The distribution of FLS scores in the F × W population was not normal (P=0.0021)(Fig. 2), but this can be expected according to the Central Limit Theorem. There was a positive skewness of 0.53, suggesting segregation is contributing to lines with more resistance to *C. sojina* Hara. The mean FLS score was relatively low at 3.19 ± 1.02, with scores ranging from 1.00-6.33. Forrest had an FLS score of 2.25 ± 0.43, and Williams 82 had an FLS score of 5.00 ± 2.73. There were 26 lines with FLS scores lower than Forrest (Table 1), and seven lines with higher FLS scores than Williams 82 (Table 2). Resistant lines had an average score of 1.70 ± 0.31 and susceptible lines had an average score of 5.94 ± 0.30.

B. Construction of Genetic Linkage Maps

A genetic map was created of the E × F population with a total of 1,959 markers across 20 chromosomes (Fig 3.). The total map length was 2,121.01 cM with an average distance between markers of 1.08 cM (Table 3). The average chromosome length was 105.05 cM with 97.95 markers on each chromosome. The largest chromosome was Chr. 19 with a length of 133.66 cM and 95 markers, while the shortest was Chr. 16 with a length of 84.27 cM and 55

markers. The most genetically dense chromosome was Chr. 3 with 1.17 markers/cM. The gaps of < 5 cM were at a rate of 99.97%.

Across 20 chromosomes in the F × W population, 2,186 were identified (Fig. 4). The map was 2,105.23 cM long with an average distance of 0.97 cM separating markers (Table 4). The largest gap between markers was 74.35 cM, with 97.16% of gaps being <5 cM. Average chromosome length was 105.26 cM. The longest chromosome was Chr. 18 at 137.47 cM long with 164 markers. The shortest chromosome was Chr.16 at 83.40 cM long with 73 markers. The most genetically dense chromosome was Chr. 20 with 1.81 markers/cM.

C. Identification of QTL

A total of three QTL were identified across both populations (Table 5). In E × F, the ss715614578–ss715615158 interval (Position: 61.81–69.27 cM) (QTL1) was identified to underlie FLS resistance on chromosome 13 (LG F). A single peak was observed at the ss715614724 marker (Position: 64.04 cM) with a LOD score of 6.36, the variation of the phenotype explained by the QTL was 14.33% (Fig. 5). The LOD threshold 4.38 was used for 95% confidence, and our QTL exceeds this criteria.

Two QTL were identified in the F × W population to underlie FLS resistance. The first is on chromosome 6 (LG C2) from ss715594329-ss715594474 (Position 87.11-99.97 cM)(QTL2). One peak was noted in this interval at ss715594440 (Position: 64.04 cM) with an LOD score of 5.16 (Fig.6). This QTL explains 5.16% of phenotypic variation. The second QTL is on chromosome 11 (LG B1) from ss715610717-ss715610843 (Position 9.90-13.04 cM) (QTL3) with a peak at ss715610720 (Position 9.94 cM) (Fig.7). This QTL explains another 6.75% of phenotypic variation. Interaction effects of the two QTL were insignificant (P=0.14). With the

95% confidence LOD threshold from the permutation test set at 4.48, we can be 95% confident that QTL2 is significant. QTL1 meets the 80% confidence threshold of 3.67.

D. Resistance Alleles

The genotypes of RILs in the E × F population that were more resistant than Essex were found to have a Forrest-like genotype at QTL1 (Table 6), whereas those that were more susceptible than Forrest to have Essex-like alleles at the same location. These results suggested that Forrest was the parent contributing to the QTL of resistance. To confirm this hypothesis, one-way ANOVA was conducted comparing FLS scores of all RILs (n=81). This test compared lines with Forrest-like alleles, Essex-like alleles, and recombinant genotypes (Fig. 8). The ANOVA test was statistically significant to 95% confidence ($F_{2,80} = 7.64$, $P < 0.0009$). Lines with Forrest-like alleles had mean FLS ratings 1.15 smaller, which equates to approximately 11.5% less foliar damage, compared to Essex-like alleles. Heterozygous lines were not statistically different from either Forrest-like or Essex-like lines.

Similar one-way ANOVAs were ran in the F × W population to see if the alleles present at the QTL of interest significantly impacted FLS score. These tests compared lines with Forrest-like alleles, Williams 82-like alleles, and recombinant alleles. The ANOVA ran on QTL2 was not significant ($F_{2,175}=2.89$, $P > 0.05$), and neither was the ANOVA ran on QTL3 ($F_{2,177}=2.38$, $P > 0.05$).

E. Gene Ontology and Kyto Encyclopedia of Genes and Genomes Pathways

Within QTL1, a wide variety of genes have been published and identified (Table 7), (Grant et al., 2010). The nearest gene to the peak at ss715614724 are the BT089187.1 and M31024.1 genes, both of which code for ribosomal protein S11. This protein resides within the

cytosolic small ribosomal subunit and plays a major role in rRNA binding and overall ribosomal structure. (UniProt Consortium, 2020).

Many genes of resistance have also been published in the QTL2 interval (Table 8). The closest known genes to the peak of the interval are AK246052.1 and AB331959.1, both of which code for the peroxisomal 3-hydroxyacyl-CoA dehydrogenase-like protein. This protein is a catalyst in fatty acid metabolic processes (UniProt Consortium 2020).

In QTL3, there are a handful of published genes (Table 9). Nearby to the interval peak are the BT094200.1 and AF004806.1 genes, which code for the 24kDa seed maturation protein. It is known that this protein resides in the endoplasmic reticulum of soybean cells, but so far its function is unknown (UniProt Consortium 2020).

Twenty-six model genes are located in the QTL1 interval that code for Leucine-rich repeat (LRR) and WRKY domain proteins (Table 10). There are 24 such model genes in QTL2 (Table 11), and six model genes in QTL3 (Table 12). All of these genes are candidate genes to code for FLS resistance in their respective intervals, as these types of genes are known to play an integral role in disease resistance (Gururani et al., 2012). Since they are genes modeled in Arabidopsis, the ways these genes work in plants are not well understood.

DISCUSSION

The parents of the E × F population were scored for FLS resistance. Forrest received an FLS score 2.3-fold higher than Essex. This confirms that Forrest is more susceptible against *C. sojina* Hara race 15 than Essex. These results aligned with those presented in prior studies on resistance to race 2 (Sharma and Lightfoot, 2017). Since our histogram fit the normal distribution, the skewness was near zero, suggesting that the segregation equally contributed to high and low FLS scores. The same was done for the parents of F × W and Williams 82 was

2.22-fold more susceptible than Forrest. This confirms that Forrest is the resistant parent for this population. The population had a negative skewness of -0.53, suggesting that segregation contributed more to low FLS scores. However, based on the Central Limit Theorem, this is not actually the case (Reeve, 2016). Since only 190 lines were used out of 1,025 original lines, a normal distribution would be expected across the whole population. It should also be noted that in the E × F experiments, Forrest received a mean score of 5.75 ± 2.49, while in the F × W experiments, it received a score of 2.25 ± 0.43. Similar differences in scores were reflected in the 'Blackhawk' and 'Lincoln' checks planted in each block. Across the E × F experiments, Lincoln had an FLS score of 3.0 ± 1.58 and Blackhawk had a score of 6.5 ± 1.63. In the F × W experiments, Lincoln had an FLS score of 2.50 ± 1.5 and Blackhawk had a score of 5.50 ± 1.5. From this, we can conclude that there was higher disease pressure in the E × F experiments. Among all three parental lines, Essex holds the most resistance, while Forrest is partially resistant, and Williams 82 is the most susceptible.

A single QTL associated with FLS resistance was identified on Chr. 13 of E × F at the ss715614578–ss715615158 interval, which coincides with the region of SNP41647 that is known for *Rcs* (PI594891) in linkage group F (Pham et al., 2015). PI594891 is a Chinese plant introduction, and its resistance pathway is not yet well documented (Hoskins, 2011). QTL1 could be allelic to *Rcs* (PI594891). It is believed that this resistance gene is conditioned by *Rcs3*, but it likely carries different resistance alleles from one or two other genes (Pham et al., 2015). Two QTL, QTL2 and QTL3, were indicated in the F × W study for association with FLS. These genes were found on Chr. 6 and Chr. 11, respectively. Prior research indicates that neither of these QTL are a part of the Rcs3 gene. QTL 2 on chromosome 6 has been reported in the E × F population in prior studies (Sharma and Lightfoot, 2014). This experiment used Race 2 of *C.*

sojina Hara, so this implies QTL2 holds resistance to multiple races. It was also found to be allelic to the *Rcs2* gene. This confirms this QTL's existence across multiple populations. QTL3 has not been priorly published.

In the E × F study, Forrest contributed the resistance allele. These results are contradictory to prior studies on race 2, in which Essex donated the resistance allele (Sharma and Lightfoot, 2017). Since *Rcs2* generally confers resistance to race 2, we assumed the existence of a different resistance mechanism for race 15. Although it seems counterintuitive for Forrest to donate the resistant allele, it might be possible since Forrest was only partially susceptible. The use of only Race 15 of *C. sojina* Hara may have also played a role in this finding. More research should be conducted on which specific races Forrest is susceptible to. It is possible Race 15 is one that Forrest holds resistance for. Many prior resistance tests use mixed races, which can skew results when individual races are used. However, since the two one-way ANOVA for F × W was insignificant, there is no simple way to determine which parent donated the allele of resistance. This could be due to lower disease pressure within the F × W greenhouse assays. Higher disease pressure would allow us to see if there are more minor QTL adding to the disease resistance in this population. Since Forrest is the shared parent in the two populations, we can hypothesize that Forrest would also donate the resistance allele in F × W. Future studies should also be conducted to see if epistatic effects are at play.

In this study, all QTL were minor. QTL1 contributed 14.33% of variance, QTL2 contributed 6.01% of variance, and QTL3 contributed 6.75% of variance in their respective populations. This is probably due to the low disease pressure across the experiments. Therefore, differences among genes of small effect might not have been identified. Future research is needed under field or greenhouse conditions with relatively high disease pressure to confirm the

presence of the QTL and identify any interaction with the environment. The use of mixed races or other individual races of *C. sojina* Hara would be also beneficial to better understand the underlying mechanism of resistance and the role of the QTL. These markers should be used in future breeding projects to fine-tune marker-assisted selection for resistance to FLS.

QTL1 was found to be associated with ribosomal S11 protein. In soybeans, it was found that ribosomal S11 was significantly elevated when immature plants were treated with 2,4 D (Gantt and Key, 1985). Since this study, the presence of S11 has been associated with cellular proliferation. It is abundant in meristematic tissue and allows the plant to produce new cells efficiently (Lenvik, 1994). To this end, we can hypothesize that the found SNP alters the amount of S11 produced in the plant and allows it to overcome damage from *C. sojina* Hara.

QTL2 is associated with peroxisomal 3-hydroxyacyl-CoA, and these peroxisomes have been documented to be connected to various cell functions in soybeans. Major functions include fatty acid β-oxidation, photorespiratory glycolate metabolism, the glyoxylate cycle, metabolite transport, and stress response (Arai et al., 2008). Further research should be done to investigate this connection between peroxisomes and FLS resistance.

While the exact function of the 24kDa seed maturation protein associated with QTL3 is unknown, there is prior research on its expression in plants during the maturation process. This protein was able to be detected in the final stages of seed maturation in the parenchyma and aleurone layers of the seed coat. The gene coding for 24kDa seed maturation protein was also well expressed in vegetative tissues that had been wounded by pathogens, suggesting it also plays a part in wound response (Dhaubhadel et al., 2005). Future studies should be conducted to solidify this link, along with experiments to confirm the link between the model genes reported and FLS resistance.

CONCLUSIONS

In summary, we report a total of three QTL associated with FLS resistance. The QTL in E × F is related to Rcs (PI594891) and production of the S11 ribosomal protein that aids in cell proliferation. The associated markers should be used in future projects to stack resistance genes for FLS. Two novel QTL were reported in the F × W population on Chr. 6 and Chr. 11 that are associated with the production of peroxisomal 3-hydroxyacyl-CoA dehydrogenase-like protein and 24kDa seed maturation protein, respectively. Both proteins are associated with wounding response and could prove useful for future breeding projects aiming at FLS resistance. QTL2 on Chr.6 was also confirmed in prior experiments, suggesting it holds resistance to multiple races of *C. sojina* Hara. Environment played a large part in our experiments, and future studies should be conducted with higher and more consistent disease pressure to determine if the identified QTL could confer a higher percentage of resistance. Overall, Forrest and its derivatives are a good source for the advancement of FLS resistance in soybean.

TABLES

Table 1: Lines with FLS scores less than Forrest (2.25).

Line	FLS Score
F × W 30	2.00
F × W 50	1.75
F × W 52	2.00
F × W 60	2.00
F × W 90	1.00
F × W 108	1.00
F × W 125	1.75
F × W 132	2.00
F × W 151	1.75
F × W 153	2 .00
F × W 157	1.25
F × W 158	2.00
F × W 173	2.00
F × W 188	1.5
F × W 191	1.75
F × W 192	1.5
F × W 200	1.25
F × W 201	2.00
F × W 205	2.00
F × W 215	2.00
F × W 219	1.5
F × W 249	2.00
F × W 263	1.75
F × W 266	1.67
F × W 269	1.75
F × W 282	1.5

Table 2: Lines with FLS Scores higher than Williams 82 (5.00).

Line	FLS Score
F × W 121	5.5
F × W 129	6.33
F × W 147	6.00
F × W 176	6.00
F × W 178	6.00
F × W 227	5.5
F × W 243	6.25

Table 3: Characteristics of genetic map of E × F population.

Chromosome	**Number of Markers**	**Genetic distance (cM)**	**Average distance between markers (cM)**	**Gaps ≤ 5 (%)**	**Maximum gap (cM)**
1	76	126.05	1.68	90.78	26.33
2	81	115.46	1.44	97.53	30.13
3	128	108.79	0.85	98.43	16.48
4	109	98.64	0.91	99.08	9.90
5	94	94.61	1.01	96.80	53.42
6	128	114.33	0.90	98.43	41.76
7	90	101.99	1.14	96.66	19.76
8	98	99.97	1.03	96.93	45.04
9	76	96.45	1.28	96.05	24.39
10	91	114.06	1.26	95.60	22.35
11	87	89.45	1.04	95.49	13.28
12	80	91.25	1.15	97.50	33.12
13	163	94.47	0.58	99.38	6.07
14	85	110.02	1.30	95.29	18.32
15	113	114.66	1.02	99.11	71.47
16	55	84.27	1.56	98.18	55.73
17	84	94.97	1.14	95.23	24.60
18	150	133.35	0.89	98.66	13.95
19	76	133.66	1.51	94.73	42.00
20	95	104.56	1.11	94.73	17.67
Total	1959	2121.01	1.08	99.97	71.47

Table 4: Characteristics of the F × W genetic linkage map

Chromosome	***Number of Markers***	***Genetic distance (cM)***	***Average distance between markers (cM)***	***Gaps ≤ 5 (%)***	***Maximum gap (cM)***
1	105	121.60	1.15	99.04	69.88
2	155	117.21	0.75	58.06	8.61
3	109	108.05	0.99	98.16	32.15
4	62	112.24	1.81	96.77	74.35
5	97	92.39	0.95	94.84	10.23
6	116	114.33	0.95	95.87	10.18
7	123	98.00	0.79	98.27	6.10
8	151	100.39	0.66	98.01	7.06
9	123	93.27	0.75	99.18	10.84
10	110	114.12	1.03	98.18	57.99
11	86	88.67	1.03	95.34	12.01
12	72	88.77	1.03	91.66	11.52
13	152	91.36	1.26	98.68	8.70
14	61	111.40	1.82	91.80	27.86
15	107	115.74	1.08	96.26	14.96
16	73	83.40	1.14	97.26	48.07
17	101	94.97	0.94	98.01	6.86
18	164	137.47	0.83	98.17	39.22
19	125	115.39	0.92	96.80	10.9
20	94	106.46	1.13	94.68	28.92
Total	2186	2105.23	0.97	97.16	74.35

Table 5: Summary table of all QTL reported.

Name	Population	Interval	LG/Chr	Position of Interval (cM)	Position (cM)	LOD	R^2 (%)
QTL1	E × F	ss715614578 – ss715615158	F/13	61.81–69.27	64.04 (ss715614724)	6.64	14.33
QTL2	F × W	ss715594329 - ss715594474	C2/6	87.11-99.97	97.72 (ss715594440)	5.16	6.01
QTL3	F × W	ss715610717 - ss715610843	B1/11	9.90-13.04	9.94 (ss715610720)	3.39	6.75

Table 6: Genotyping results at QTL1.

Line	FLS Score	Genotype at ss715614724
E × F 2	1	C
E × F 9	1	C
E × F 10	1	C
E × F 11	1	C
E × F 29	7.25	T
E × F 54	1	C
E × F 63	6	T
Essex	1.5	T
Forrest	5.75	C

Table 7: All published genes in the ss715614578-ss715615158 (QTL1) interval on chromosome 13.

Gene	Protein
BT096972.1	ABC transporter/family member 1-like
FJ014823.1	protein kinase
BT093809.1	calmodulin-like protein 5-like
FJ014792.1	calmodulin-binding receptor-like cytoplasmic kinase
GQ422779.1	bifunctional purple acid phosphatase 26-like
BT089187.1*	ribosomal protein S11
M31024.1*	ribosomal protein S11
BT097035.1	pre-rRNA-processing protein TSR2 homolog
BT097614.1	CASP-like protein N24-like
BT094321.1	formate dehydrogenase 1, mitochondrial-like
DQ468343.1	SNI1
CYP93C1v2	cytochrome P450 monooxygenase CYP3C1v2p
ifs2	isoflavone synthase 2
CYP93C1	isoflavone synthase 2
FJ014793.1	receptor-like protein kinase HSL1-like
KC876033.1	Drought-induced family protein
BT089855.1	17.5 kDa class I heat shock protein-like
AK244336.1	mediator-associated protein 2-like
BT099462.1	mediator-associated protein 2-like
DQ857259.1	Dof9
BT096749.1	40S ribosomal protein S6-like
BT094501.1	probable RNA 3'-terminal phosphate cyclase-like protein-like
BT097216.1	epoxide hydrolase 2-like
BT098969.1	monoglyceride lipase-like
AK285956.1	secretory carrier-associated membrane protein-like
BT094395.1	secretory carrier-associated membrane protein-like
BT095720.1	putative 12-oxophytodienoate reductase 11-like

*Closest genes to interval peak

Table 8: Published genes in the ss715594329-ss715594474 (QTL2) interval on chromosome 6.

Gene	Protein
BT097968.1	Early nodulin-like protein 2-like
AK246086.1	NAC domain protein NAC3
NAC3	NAC domain protein NAC3
SNAC23	NAC transcription factor
SNAC41	NAC domain-containing protein 18-like
AK286350.1	Proteasome IOTA subunit
IOTA	Proteasome IOTA subunit
AK246052.1*	Peroxisomal 3-hydroxyacyl-CoA dehydrogenase-like protein
AB331959.1*	Peroxisomal 3-hydroxyacyl-CoA dehydrogenase-like protein

*Closest genes to interval peak

Table 9: Published genes in the ss715610717-ss715610843 (QTL3) interval on chromosome 11.

Gene	Protein
BT094200.1*	24kDa seed maturation protein
AF004806.1*	24kDa seed maturation protein
BT094706.1	Polyneuridine-aldehyde esterase-like
CYP82C1	Cytochrome P450 CYP82C1
BT093995.1	NASP-related protein sim3-like
AK286723.1	Oxygen-evolving enhancer protein 1%2C chloroplastic-like
BT095172.1	Oxygen-evolving enhancer protein 1%2C chloroplastic-like
Y10493.1	Putative cytochrome P450
bZIP118	bZIP transcription factor bZIP118
GBF1	G-box binding factor
BT094253.1	HVA22-like protein k-like
BT094413.1	Tyrosyl-tRNA synthetase-like

*Closest genes to interval peak

Table 10: Model genes in the QTL1 interval.

Gene	Pfam (Family)	Panther (Function)
Glyma13g24330	PF00069(Protein kinase domain)	PTHR24420 (Leucine-Rich Repeat Receptor-Like Protein Kinase)
Glyma13g24340*	PF00560(Leucine Rich Repeat)	PTHR24420 (Leucine-Rich Repeat Receptor-Like Protein Kinase)
Glyma13g24550	PF00560 (Leucine Rich Repeat)	PTHR24420 (Leucine-Rich Repeat Receptor-Like Protein Kinase)
Glyma13g24980	PF07714 (Protein tyrosine kinase)	PTHR24420 (Leucine-Rich Repeat Receptor-Like Protein Kinase)
Glyma13g25340	PF07714 (Protein tyrosine kinase)	PTHR24420 (Leucine-Rich Repeat Receptor-Like Protein Kinase)
Glyma13g25420	PF00931 (NB-ARC domain)	PTHR23155 (Leucine-Rich Repeat-Containing Protein)
Glyma13g25440	PF00931 (NB-ARC domain)	PTHR23155 (Leucine-Rich Repeat-Containing Protein)
Glyma13g25724	PF07714 (Protein tyrosine kinase)	PTHR24420 (Leucine-Rich Repeat Receptor-Like Protein Kinase)
Glyma13g25750	PF00931 (NB-ARC domain)	PTHR23155 (Leucine-Rich Repeat-Containing Protein
Glyma13g25780	PF00560 (Leucine Rich Repeat	PTHR23155 (Leucine-Rich Repeat-Containing Protein)
Glyma13g25800	PF11883 (Domain of unknown function (DUF3403))	PTHR24420 (Leucine-Rich Repeat Receptor-Like Protein Kinase)
Glyma13g25811	PF07714 (Protein tyrosine kinase)	PTHR24420:SF430 (Leucine-Rich Receptor-Like Protein Kinase)
Glyma13g25820	PF07714 (Protein tyrosine kinase)	PTHR24420 (Leucine-Rich Repeat Receptor-Like Protein Kinase)
Glyma13g25920	PF00931 (NB-ARC domain)	PTHR23155 (Leucine-Rich Repeat-Containing Protein)
Glyma13g25950	PF00560 (Leucine Rich Repeat)	PTHR23155:SF121
Glyma13g25970	PF00931(NB-ARC domain)	PTHR23155 (Leucine-Rich Repeat-Containing Protein).
Glyma13g26141	PF00931 (NB-ARC domain)	PTHR23155 (Leucine-Rich Repeat-Containing Protein)
Glyma13g26230	PF00931 (NB-ARC domain)	PTHR23155 (Leucine-Rich Repeat-Containing Protein)
Glyma13g26310	PF00560 (Leucine Rich Repeat)	PTHR23155 (Leucine-Rich Repeat-Containing Protein)
Glyma13g26380	PF00931 (NB-ARC domain)	PTHR23155 (Leucine-Rich Repeat-Containing Protein)
Glyma13g26400	PF01582 (TIR domain)	PTHR11017 (Leucine-Rich Repeat-Containing Protein)

Glyma13g26420	PF00931 (NB-ARC domain)	PTHR11017 (Leucine-Rich Repeat-Containing Protein)
Glyma13g26450	PF01582 (TIR domain)	PTHR11017 (Leucine-Rich Repeat-Containing Protein)
Glyma13g26460	PF01582 (TIR domain)	PTHR11017 (Leucine-Rich Repeat-Containing Protein)
Glyma13g26530	PF00560 (Leucine Rich Repeat)	PTHR23155 (Leucine-Rich Repeat-Containing Protein)
Glyma13g26650	PF01582 (TIR domain)	PTHR11017:SF20 (Subfamily Not Named)

Table 11: Model genes in within the QTL2 interval.

Gene	Pfam (Family)	Panther (Function)
Glyma06g37100	PF03106 (WRKY DNA - binding domain)	-
Glyma06g37441	PF07714 (Protein tyrosine kinase)	PTHR24420:SF692
Glyma06g37505	PF07714 (Protein tyrosine kinase)	PTHR24420:SF692
Glyma06g39725*	PF00931 (NB-ARC domain)	PTHR23155 (Leucine-Rich Repeat-Containing Protein)
Glyma06g39930*	PF00954 (S-locus glycoprotein family)	PTHR24420:SF703
Glyma06g39943	PF01582 (TIR domain)	PTHR23155 (Leucine-Rich Repeat-Containing Protein)
Glyma06g39990	PF00931 (NB-ARC domain)	PTHR23155 (Leucine-Rich Repeat-Containing Protein)
Glyma06g40000	PF00069 (Protein kinase domain)	PTHR24420:SF703
Glyma06g40021	PF07714 (Protein tyrosine kinase)	PTHR24420:SF430 (Leucine-Rich Receptor-Like Protein Kinase)
Glyma06g40030	PF00954 (S-locus glycoprotein family)	PTHR24420 (Leucine-Rich Repeat Receptor-Like Protein Kinase)
Glyma06g40050	PF00954 (S-locus glycoprotein family)	PTHR24420 (Leucine-Rich Repeat Receptor-Like Protein Kinase)
Glyma06g40110	PF11883 (Domain of unknown function (DUF3403)	PTHR24420 (Leucine-Rich Repeat Receptor-Like Protein Kinase)
Glyma06g40130	PF00069 (Protein kinase domain)	PTHR24420:SF703
Glyma06g40141	PF00069 (Protein kinase domain)	PTHR24420 (Leucine-Rich Repeat Receptor-Like Protein Kinase)
Glyma06g40161	PF07714 (Protein tyrosine kinase)	PTHR24420:SF430 (Leucine-Rich Receptor-Like Protein Kinase)
Glyma06g40170	PF00954 (S-locus glycoprotein family)	PTHR24420:SF703
Glyma06g40240	PF00954 (S-locus glycoprotein family)	PTHR24420 (Leucine-Rich Repeat Receptor-Like Protein Kinase)
Glyma06g40350	PF00954 (S-locus glycoprotein family)	PTHR24420:SF703

Glyma06g40370	PF07714 (Protein tyrosine kinase)	PTHR24420 (Leucine-Rich Repeat Receptor-Like Protein Kinase)
Glyma06g40400	PF00954 (S-locus glycoprotein family)	PTHR24420:SF432
Glyma06g40461	PF00069 (Protein kinase domain)	PTHR24420:SF430 (Leucine-Rich Receptor-Like Protein Kinase)
Glyma06g40480	PF00954 (S-locus glycoprotein family)	PTHR24420 (Leucine-Rich Repeat Receptor-Like Protein Kinase)
Glyma06g40490	PF00954 (S-locus glycoprotein family)	PTHR24420 (Leucine-Rich Repeat Receptor-Like Protein Kinase)
Glyma06g40515	PF00954 (S-locus glycoprotein family)	PTHR24420 (Leucine-Rich Repeat Receptor-Like Protein Kinase)

*Closest genes to interval peak

Table 12: Model genes in the QTL3 interval.

Gene	Pfam (Family)	Panther (Function)
Glyma11g06260*	PF05659 (Arabidopsis broad-spectrum mildew resistance protein RPW8)	PTHR23155 (Leucine-Rich Repeat-Containing Protein)
Glyma11g06270	PF00560 (Leucine Rich Repeat)	PTHR23155 (Leucine-Rich Repeat-Containing Protein)
Glyma11g06451	PF00069 (Protein kinase domain)	PTHR24420 (Leucine-Rich Repeat Receptor-Like Protein Kinase)
Glyma11g06740	PF00069 (Protein kinase domain)	PTHR24420 (Leucine-Rich Repeat Receptor-Like Protein Kinase)
Glyma11g06750	PF00069 (Protein kinase domain)	PTHR24420 (Leucine-Rich Repeat Receptor-Like Protein Kinase)
Glyma11g07175	PF07714 (Protein tyrosine kinase)	PTHR24420:SF822

*Closest genes to interval peak

FIGURES

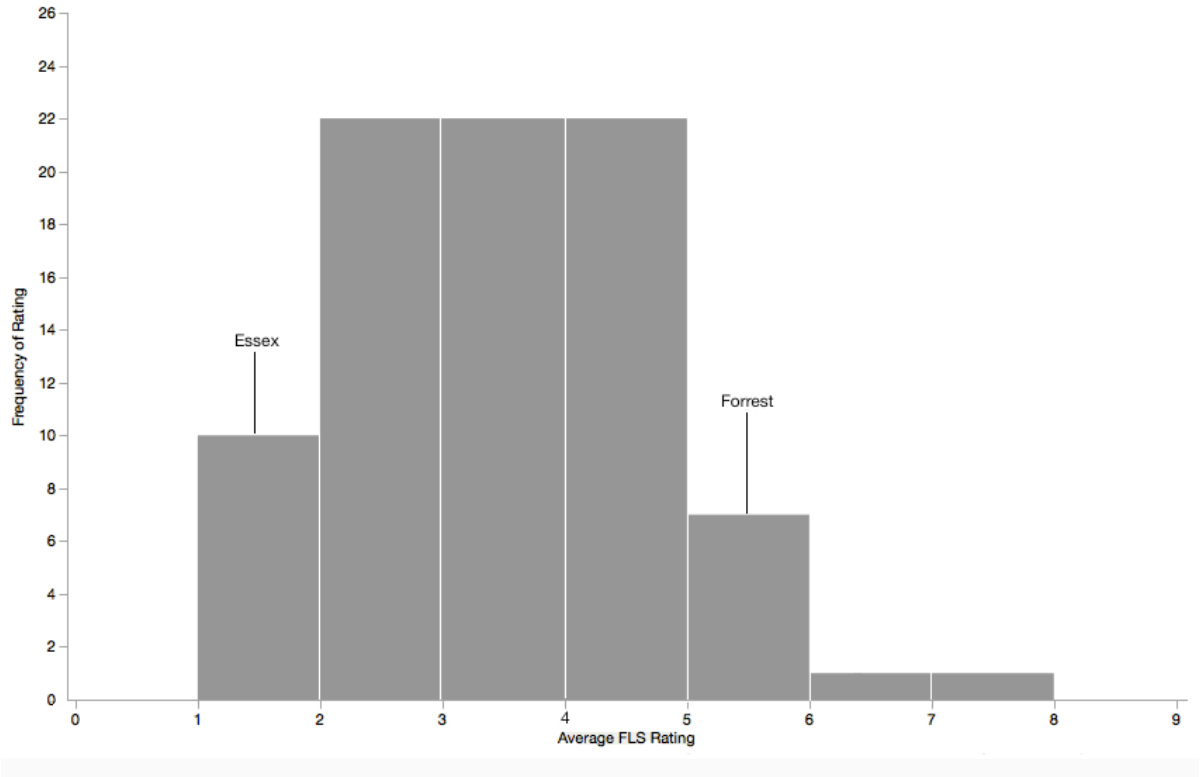


Figure 1: Histogram depicting the frequency of FLS scores in the E × F experiments.

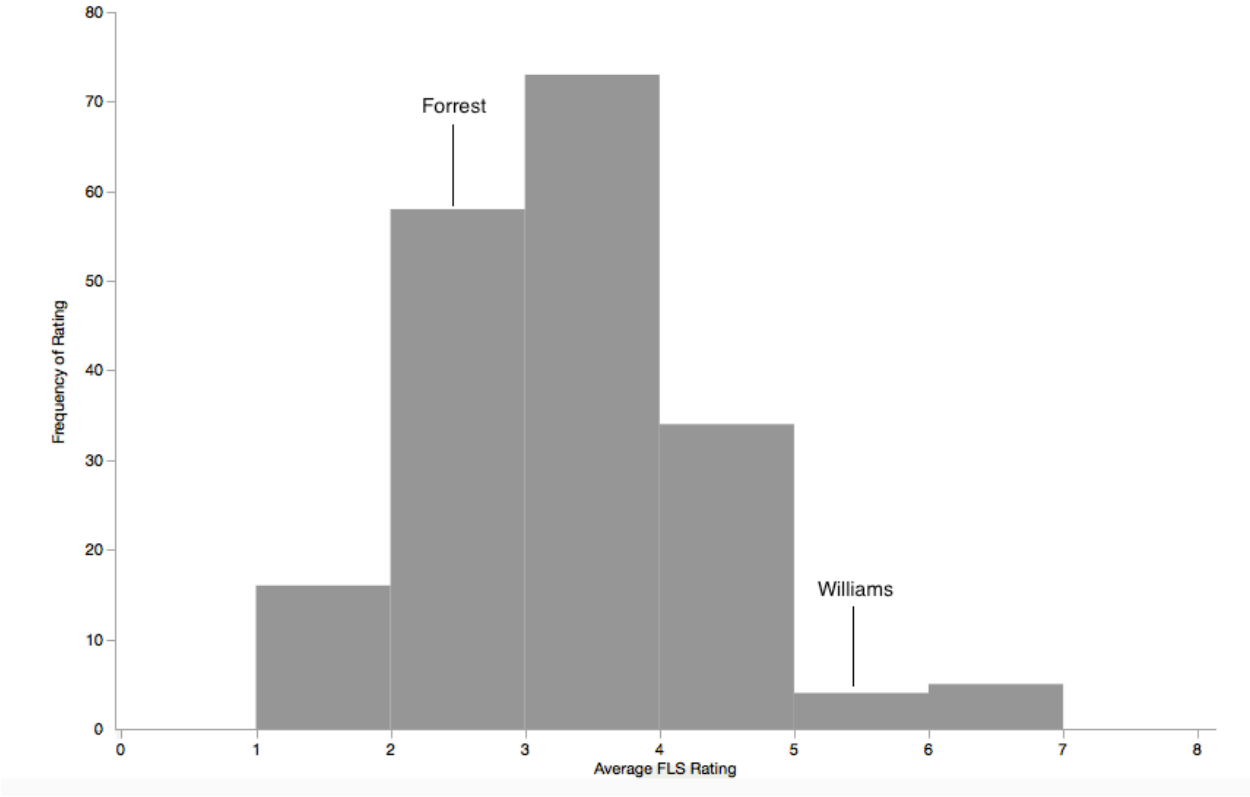


Figure 2: Histogram depicting the frequency of FLS scores across the F × W experiments.

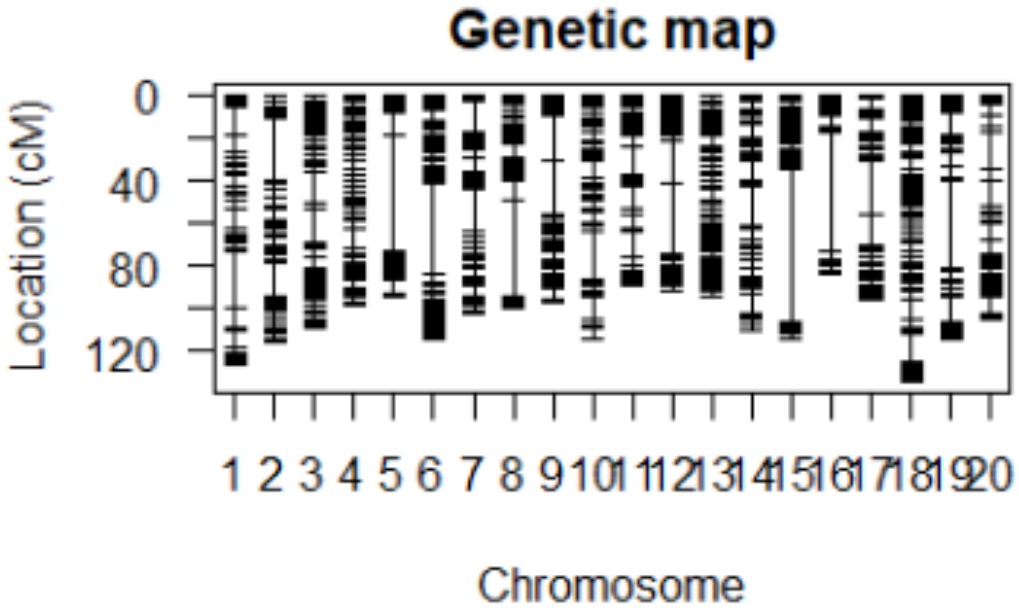


Figure 3: Genetic linkage map of E × F population.

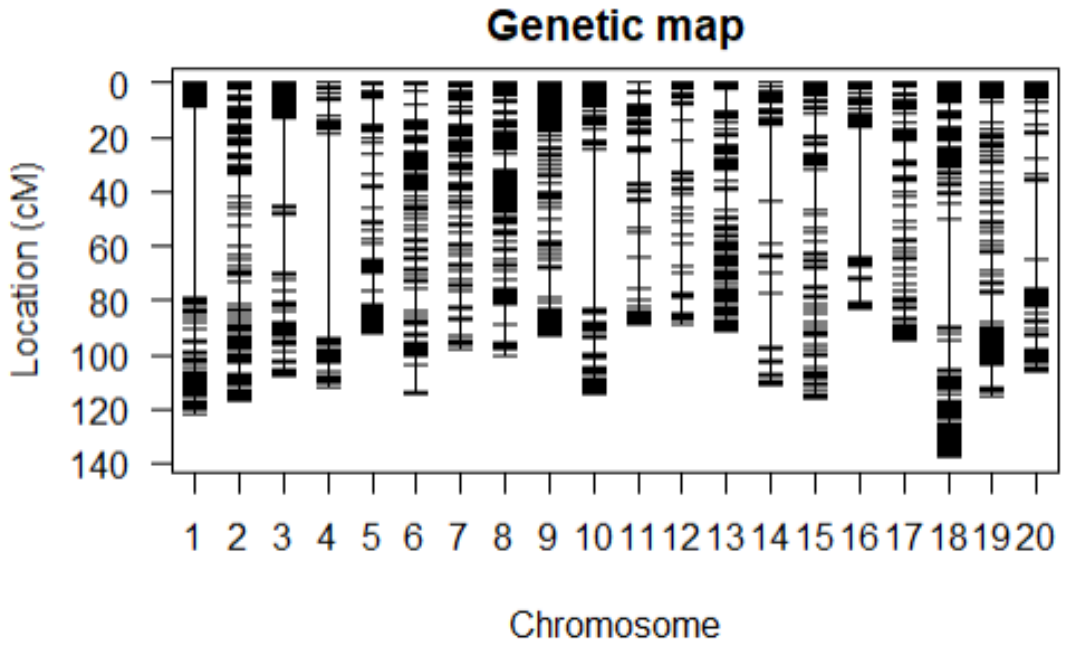


Figure 4: Genetic linkage map of F × W population

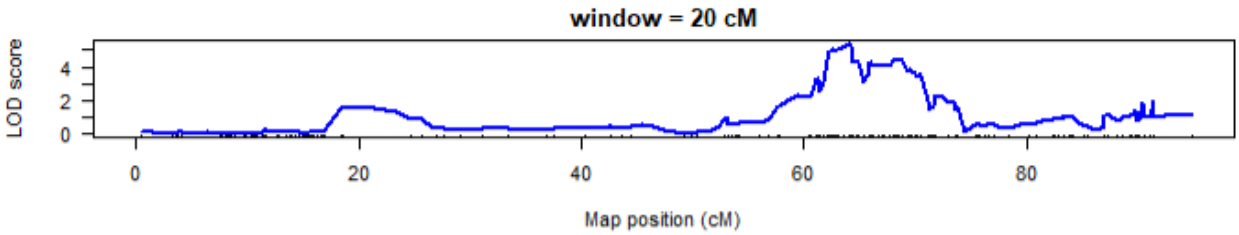


Figure 5: CIM interval of QTL1.

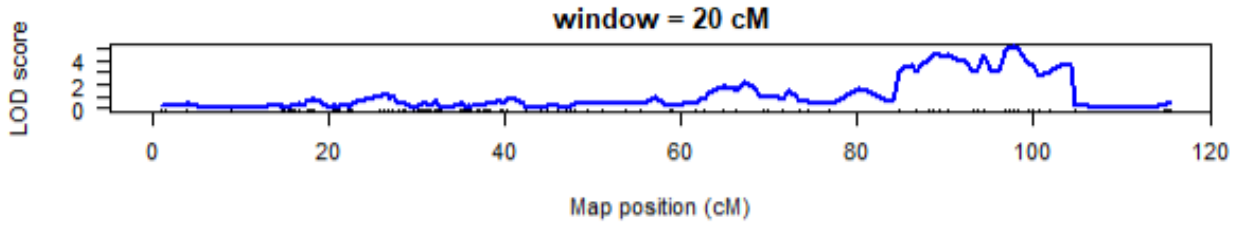


Figure 6: CIM interval of QTL2.

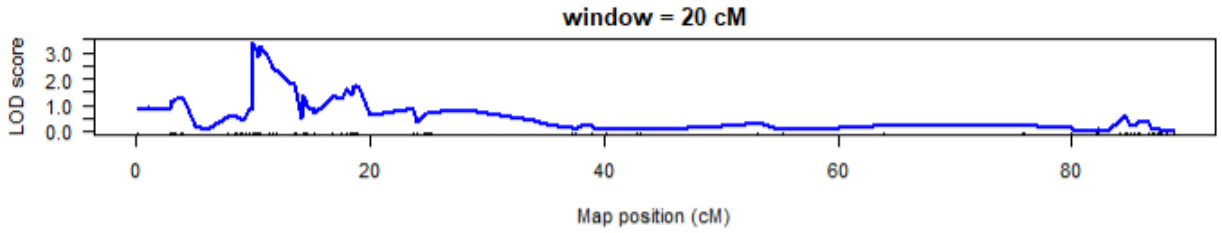
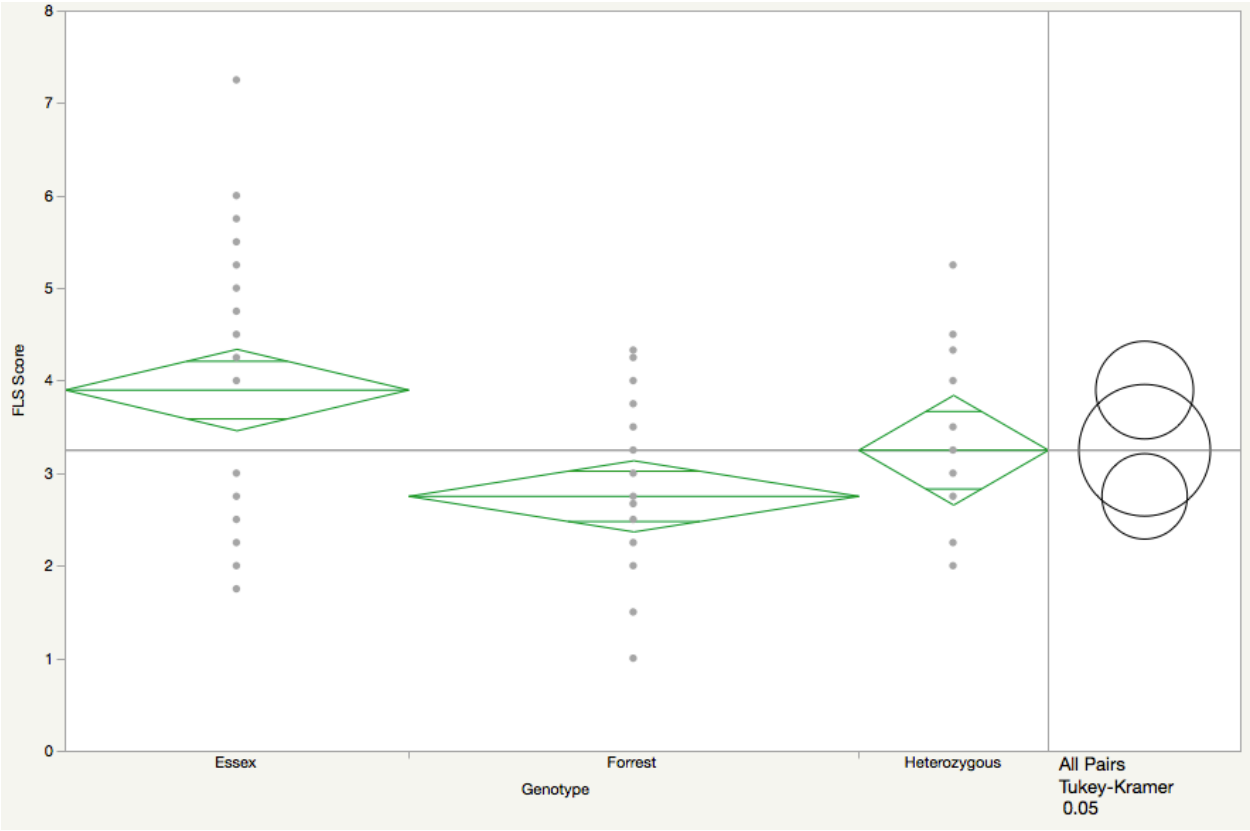


Figure 7: CIM interval of QTL 3.



Level	Grouping (Allele)	Mean
Essex	T	3.90
Recombinant	TC	3.25
Forrest	C	2.75

Figure 8: One-way ANOVA comparing genotypes at QTL1 ($F_{2,80}$=7.64, P<0.0009).

CHAPTER 3

ANALYZING AGRONOMIC TRAITS OF TWO SOYBEAN POPULATIONS UNDER FIELD CONDITIONS

ABSTRACT

Soybeans are an important cash crop globally that majorly contributes to food, fiber, and oil production. Many variables contribute to a farmer's decision about which soybean variety to plant, some of which include high germination, maturity group, and perspective yield. Secondary considerations can include resistance to a specific pathogen or pest that is prevalent in the area. For this reason, breeders need to take various agronomic traits into consideration whenever developing new soybean lines, as needs and desires can vary across geographic regions. To this end, we have characterized the germination rate, flower color, days to 90% flowering, days to 90% maturity, and yield of the 'Essex' × 'Forrest' population and the 'Forrest' × 'Williams' population. These populations have priorly been characterized for their resistance to *Cercospora sojina*, and after understanding the agronomic characteristics, can be used for implementation of disease resistance. One line, F×W125, stood out among all others for desired agronomic traits and Frogeye Leaf Spot resistance. This line should be used for future breeding projects to develop the latest elite lines.

Keywords: soybeans, C.sojina, flowering, maturity, agronomic traits

INTRODUCTION

Soybeans are a major crop globally, with production in the U.S. climbing every year (Pagano & Miransari, 2016). It is commonly used for oil, feed, and biofuels, and accounts for 90% of U.S. oilseed production. More than 80% of U.S. soybeans are grown in the midwestern states, where it is rotated annually with corn to add nitrogen back to the soil. The United States

government subsidizes this crop, making market fluctuations less volatile. For all these reasons, it is a safe plant for many farmers to choose, and production has only increased every year since 1990 (Ash, Livezey, & Dohlman, 2006).

There are many factors farmers think about when selecting which soybean variety to plant each year. One of the most important is maturity group (MG), which indicates how long it will take for a variety to reach full pod set. The larger the number, the longer the variety takes to fully flower and mature. In Southern Illinois, many farmers use MG 4-5. The southern United States will use MG 5-6 in order to optimize the full growing season (Mourtzinis & Conley, 2017). Many times, farmers want shorter maturity plants for added flexibility when dealing with rainy planting seasons or early snow. These are all considerations to make when farmers are selecting lines and as breeders work to make new varieties.

Flowering and Maturity are largely attributed to three main genes: *E1, E2,* and *E3* (Bernard, 1971; Watanabe et al., 2009). Soybeans are a short-day photoperiod sensitive plant that induces flowering when there is a <12 hr day length (Destro, Carpentieri-Pipolo, Kihl, & Almeida, 2001). Multiple quantitative trait loci (QTL) have been associated with photoperiod responses and the correlating *E1*, *E2*, and *E3* genes (Wantanabe, Harada, & Abe, 2012). Both traits are also tangibly intertwined with plant height and architecture, which is controlled by the Dt_1 gene. (Bernard, 1972). There are two different methods of soybean growth: determinate (lateral growth halts after flowering), or indeterminate (continue lateral growth after flowering). If flowering is induced early in a determinate line, total pod number can be drastically decreased, along with yield (Bernard, 1972). It is understood in the soybean breeding community that there is no one gene that correlates with yield, but instead it is a mixture of hundreds of various genes that contribute to the overall well-being and success of the plant.

Traditionally bred soybeans are an equal blend of both parents. For this reason, many breeders focus on crossing two elite lines to try and achieve a hybrid progeny with better characteristics than either parent. To implement novel genes, backcrossing would be used over multiple generations to try and implement a new trait of interest while maintaining all other characteristics from an elite parent (Concibido et al., 2003). A line with a plethora of undesirable traits increases the odds of linkage drag into progeny lines and should be avoided. To this end, the objectives of this experiment were to characterize two soybean RIL populations for their desirable agronomic traits, and identify which lines carry the least linkage drag for Frogeye Leaf Spot (FLS) resistance.

MATERIALS AND METHODS

A. Plant Material

The Essex x Forrest (E × F) and Forrest x Williams (F × W) RIL populations were used for this field experiment, as are described in Chapter 1. These lines were allowed to self-pollinated in the field to provide refreshed seed for the next year's field experiments and other lab research.

B. Field Design and Conditions

Field experiments took place at the Agronomy Research Center at Southern Illinois University Carbondale. There were two years of experiments, with each year taking place in a different location on the research center. Each year was divided into two blocks with the F × W and E × F populations inside (Fig. 9). A complete randomized block design was utilized. Each line was planted in two-row plots 10 ft in length. Four foot gaps separated individual plots. Three hundred seeds were planted in each plot with 150 seeds in each row.

No irrigation was used and plants were allowed to experience natural field conditions. Summer 2019 was particularly rainy during the planting season, which prevented planting until May 28. After the field had been planted, Carbondale experienced a drought that likely contributed to decreased germination. As the summer progressed, there were pest issues. Deer ate the tops of many different lines. Summer 2020 had more ideal conditions, but planting was delayed due to the COVID-19 pandemic and logistical issues. The field was planted on June 2, and there was light rainfall in the following weeks.

Both fields were soil tested in the spring before planting, and no fertilizer was applied in accordance with the results. Pre-emergence herbicides used were FirstRate (cloransulam-methyl) at a rate of .60 oz/A and Prefix (S-metolachlor, sodium salt of fomesafen) at a rate of 40 oz/A. Post-emergence herbicides applied were Select (clethodim) at a rate of 4.8 oz/A, Flexstar (sodium salt of fomesafen) at a rate of .5 pts/A, Dual (S-metolachlor) at a rate of 9.6 oz/A, and FirstRate (cloransulam-methyl) at a rate of .12 oz/A.

C. Note Taking

Five different notes were taken during the growing season: germination, flower color, days to 90% flowering, days to 90% maturity, and yield. Germination notes were taken two weeks after planting to allow for slow-germinating lines to sprout. The field was walked twice a week post-germination to monitor growth stages of the plants and any pest infestations. Days to 90% flowering was recorded when 90% of the flower buds in a given plot had opened. When a plot reached this milestone, the date was recorded and days to 90% maturity was calculated. This method was also used for days to 90% maturity. This note was taken when 90% of the plants in a plot were fully dried down and ready for harvest. The plots were harvested with a two-row small plot combine and individually bagged. After harvest was complete, all plot bags were cleaned

using a shaking gravity table. This clean seed was weighed in grams and was converted into kg/ha for analysis.

D. Statistical Analysis

JMP Pro 15 software was used for all distributions, means, t-tests, and other statistical analyses. A nested ANOVA model was used to measure blocking interactions using the Fit Model function with blocks nested within the year of experiment. A 95% confidence interval was used for all significance tests.

RESULTS

A. Germination

In 2019, 250 F × W lines were planted in the field experiment. This seed stock had not been planted the season prior, so it had lower germination. A total of 154 of the lines germinated, and a germination test was conducted to determine the best lines to plant the following year. In 2020, all lines that germinated in 2019 were replanted, along with an extra 50 lines for a total of 204 lines. 172 of these lines germinated and grew to maturity under field conditions in 2020. In 2019, there was a germination rate of 61.6%. The 2020 growing season had a germination rate of 84.31%. Out of all the lines planted across the two growing seasons, 71% of lines germinated.

The 91 E × F populations were planted in 2019, and 85 lines germinated and continued to maturity, for a germination rate of 93%. These 85 lines were replanted in 2020, and 74 germinated and provided seed for future projects. The 2020 growing season had a germination rate of 88%. This leaves 81% of the original lines for future experiments.

B. Flower Color

The F × W population maintained all white flowers through both growing seasons. The E × F population, however, continued to segregate by flower color (Fig.10). In 2019, 45 lines had

white flowers, 38 lines had purple flowers, and 2 lines had a mix of purple and white. In 2020, 32 lines had white flowers, 30 lines had purple flowers, and 12 had a mix of purple and white flowers. Mixed flower colors increased to 16.21% of the population in 2020 from a mere 2.35% in 2019.

C. Days to 90% Flowering

The nested ANOVA ran on the F × W population determined that there was not a significant blocking interaction ($P < 0.05$). However, there was a significant difference between lines ($F_{189,462} = 1.60$; $P < 0.0001$). A Tukey's HSD test distinguished two lines of interest. F×W180 took significantly longer to reach 90% flowering at 71.25 ± 1.45 days. F × W235 had a significantly shorter time to reach flowering at 61.50 ± 1.45 days. All other lines were not significantly different from one another. The mean time to 90% flowering across the F×W population was 65.69 ± 3.16 days. The distribution of days to 90% was not normal (GOF <0.0001)(Fig.11), but this can be explained with the Central Limit Theorem (Reeve, 2016). There was a positive skewness of 0.83, suggesting segregation has contributed to shorter flowering times (n=652).

The E × F population was also determined to have an insignificant blocking interaction ($P < 0.05$). There was also an no significant differences across lines ($P < 0.05$). The population had an average of 67.7 ± 4.18 days to 90% flowering. The distribution was not normal (GOF < 0.001)(Fig.12), which could be due to the missing lines that did not germinate. Like F×W, the E×F population has a positive skewness of 0.79, suggesting the population is segregating for shorter flowering periods (n=304).

An ANOVA was ran to determine if the two populations were significantly different from one another, and it indicated a significant difference ($F_{1,954} = 19.96$; $P < 0.0001$). A

student's t-test indicated that with 99.9% confidence, F×W has shorter time to 90% flowering, with a difference of 2.07 days (Table 13).

D. Days to 90% Maturity

A nested ANOVA for the F × W population indicated no significant blocking interaction ($P < 0.05$). The ANOVA indicates a significant difference in lines ($F_{186,459} = 1.53$; $P < 0.002$). However, Tukey's HSD does not distinguish a difference among lines, so no conclusions can be drawn about which lines are significant. The population overall has a mean days to 90% maturity of 135.51 ± 4.01 days. The distribution is not normal (GOF <0.001)(Fig.13), and has a negative skewness of -0.34 (n=648). This would suggest the population is segregating for longer time to maturity.

For E × F, there is a significant blocking interaction ($F_{2,213} = 5.29$, $P < 0.0057$). A Tukey's HSD test indicates that both blocks in 2019 are significantly different from the 2020 blocks (Table 14). However, there is no significant difference across lines. In 2019 Block 1, the mean days to maturity is 138.73 ± 4.38 days. 2019 Block 2 has a mean of 140.91 ± 2.32. In 2020, the overall mean days to maturity was 135 ± 0. An analysis of the distribution across all blocks shows an overall mean of 137.39 ± 3.54. The distribution is not normal (GOF <0.0001)(Fig.14), and has a positive skewness of 0.36. This suggests the population is segregating for shorter days to 90% maturity (n=295).

A student's t-test was used to compare F × W and E × F on a population level, and there is a significant difference between the two. With 99.9% confidence, we can state that E×F takes longer to reach 90% maturity, with a difference of 1.87 days (Table 13).

E. Yield

The yield distribution of F × W is normal (GOF > 0.05)(Fig.15), with a slight positive skewness of 0.55. This would indicate that segregation is contributing to lower yield (n=654). An ANOVA ran on the population shows no blocking interaction or significant difference in lines. Mean yield is 1,397.66 ± 1,185.22 kg/ha.

The E × F population also did have a normal distribution (GOF > 0.05)(Fig.16). It has a small positive skewness of 0.07, and has a bimodal pattern with peaks around 0-250 kg/ha and 2000-2250 kg/ha (n=293). This could be due to environmental factors in specific parts of the fields in both years that create more low yielding lines. The overall mean is 1,577.00 ± 1,024.23 kg/ha.

A pooled t-test was ran to compare the F×W population and the E × F population. With 95% confidence, we can state that E × F is a higher-yielding population with an average difference of 179.64 kg/ha ($P < 0.0125$) (Table 13).

F. Characteristics of FLS-resistant Lines

Similar nested ANOVAs were conducted on the lines distinguished in Chapter 1 as being resistant to FLS. There are 31 total lines that are characterized as being FLS resistant, with five of them from the E × F population and 26 of them from the F × W population. Their days to 90% flowering, days to 90% maturity, and yield were analyzed separately to distinguish which lines would be most ideal for breeding FLS resistance into current lines without linkage drag (Table 15).

A distribution analysis of the resistant lines was not normal for 90% flowering (GOF <0.0001)(Fig.17). There is a positive skewness of 1.14, implying more lines have a smaller number of days to 90% flowering (n=90). There is no significant blocking interaction in resistant

lines when looking at 90% flowering (P > 0.05). However, there is a significant difference among lines (P = 0.0041). There is a range of 18 days, from 77-64 days with a mean of 65.54 ± 3.06 days. A Tukey's HSD distinguished one line, E×F 10, that took significantly longer to reach 90% flowering. Fourteen lines took significantly less time to reach 90% flowering, with least square means ranging from 65.25-62 days. These lines were F × W 52, F × W 188, F × W 205, F × W 215, F × W 30, F × W 125, F × W 132, F × W 151, F × W 153, F × W 191, F × W 269, F × W 219, F × W 108, and F × W 192. The remaining 14 lines were not significantly different from any lines. Two lines had missing data.

As for days to 90% maturity, the distribution is not normal (GOF <0.0001)(Fig. 18). There is a slight negative skewness of 0.23, which implies there is almost an even amount of smaller and larger numbers of days to maturity (n=79). There is no significant blocking interaction (P > 0.05), but there is a significant difference in lines (P = 0.0046). Five lines had missing data, and days to 90% maturity ranged from 143.5-127.5 with a mean of 135.02 ± 4.59 days. A Tukey's HSD test identified one line that took significantly longer to reach 90% maturity: F × W 191 (143.50 days). The F × W 125 line had a significantly shorter time to 90% maturity, at 127.5 days. All other lines were not significantly different from one another.

A distribution analysis of FLS-resistant lines' yield was not normal (GOF <0.0001) (Fig. 19). There is a positive skewness of 0.50, indicating there are more lines with smaller yields (n=77). Yield in FLS-resistant lines had no significant blocking interaction (P > 0.05). There was also no significant difference among lines (P > 0.05). Yield measures ranged by 3,668.35 kg/ha from 21.52-3689.87 kg/ha, with a mean of 1,240.33 ± 1,184.01.

DISCUSSION

No distribution in the population-wide analyses were normal, but this can be explained by the Central Limit Theorem (Reeve, 2016). Not all available lines were planted, so it can be expected to see non-normal distributions. The germination decreases that were seen both years were likely due to weather conditions, the age of the seed in 2019, and pests. We would expect that with each given year the lines were planted, germination percentages would increase as selection pressure for well-germinating lines increases.

It appears that the F × W population's flower color is stable and genes for flower color are no longer segregating. This can be concluded since over both the 2019 and 2020 growing seasons, all lines had white flowers. However, E × F appears to have unstable genes for flower color, due to the increase in mixed flower colors. Another possibility for this variability is seed contamination. While all measures were used to keep individual lines pure, it is possible that the combine did not fully clean out during a plot and contaminated the next few plots of seed. This should become more obvious in future generations if the flower color stabilizes or continues to be mixed.

Both populations seem to be segregating for decreased time to 90% flowering. This is beneficial to breeding programs, as shorter flowering times are generally desired by farmers and producers in the Midwestern US. Since F × W has significantly shorter times to 90% flowering, it would be the ideal population to use if shortened flowering times were the main objective of the project. The F × W population is segregating for more days to 90% maturity, leaving a wider gap between flowering and maturity. The E × F population is segregating for shorter days to 90% maturity, which is also more ideal for farmers. However, F × W still has significantly shorter times to 90% maturity, so it is currently more ideal for breeding programs looking for shortened

time to maturity. With shorter times to flowering and maturity, farmers can still have a productive year if conditions in the spring are unconducive to planting. The sooner the beans are mature, the faster farmers can harvest and prevent frost or snow damage.

Since E × F has significantly higher yield, it would be beneficial to the many programs aiming for high-yielding varieties. The average yield in the US in 2019 was 47.4 bu/ac, which converts to approximately 3,187.70 kg/ha (Intel, 2020). Both populations have much lower yield, likely due to them being put under experimental field conditions. The populations had no seed treatments, fungicides, or other precautionary applications done like would happen in a standard production setting, likely contributing to lower yield. Since E × F has statistically higher yield, it would be less likely to drastically reduce yields in a breeding project for any other desired traits.

When deciding between using the F × W or E × F populations for breeding projects, it is important to know the goals of the project before deciding. If the intended goal is decreased time to flowering or maturity, F × W is the more qualified candidate. E × F is more likely to be beneficial in a project that is trying to increase soybean yield.

One line stands out among all others for both FLS resistance and other desirable field traits: F × W125. It has significantly lower days to 90% flowering and significantly lower days to 90% maturity when compared to other lines that are FLS-resistant (Table 16). Since no lines were deemed significantly higher in yield, this is not a measure we can take into account. However, F × W125 should be used as a line for implementing FLS resistance in current commercial lines in accordance with all of the ideal field characteristics it has obtained.

CONCLUSIONS

Field experiments are important for a comprehensive understanding of the characteristics of a given soybean line or population. When deciding on using the F×W or E×F populations for

breeding programs, it is important to note the objectives of the experiment to make the proper selection. F × W has more desirable characteristics for flower color, days to 90% flowering, and days to 90% maturity. E × F, however, is a higher yielding line. For integration of FLS resistance, the F × W125 line should be used for enhanced resistance to *C. sojina* with limited linkage drag. Future experiments should analyze these populations for potential QTL that can be associated with these desired traits.

TABLES

Table 13: Summary of field characteristics by population

Characteristic	F × W	E × F	Significance
Mean Days to 90% Flowering	65.69 ± 3.16	67.76 ±4.18	P < 0.0001
Mean Days to 90% Maturity	135.51 ± 4.01	137.39 ± 3.54	P < 0.0001
Mean Yield (bushels/hectare)	1397.66 ± 1185.22	1577.30 ± 1024.23	P<0.0125

Table 14: Results of Tukey's HSD for 90% maturity of the E × F population

Level	Grouping	Least Square Mean
2019 Block 2	A	138.50
2020 Block 1	AB	137.48
2020 Block 2	AB	137.48
2019 Block 1	B	136.47

Table 15: Field characteristics of FLS-resistant lines

Line	Days to 90% Flowering	Days to 90% Maturity	Yield (kg/ha)
F × W 30	64 ± 0.00	138.25 ± 3.26	2746.95 ± 942.92
F × W 50	65.5 ± 2.59	135 ± 6.00	32.28 ± 10.76
F × W 52	65.25 ± 1.29	135.5 ± 0.50	1639.33 ± 1512.95
F × W 60	65.5 ± 2.59	136.75 ± 2.48	1468.19 ± 1020.22
F × W 90	67 ± 3.00	129 ± 0.00	21.52 ± 0.00
F × W 108	62.75 ± 2.16	131.25 ± 3.89	1252.91 ± 743.57
F × W 125	64 ± 0.00	127.5 ± 1.50	43.05 ± 0.00
F × W 132	64 ± 0.00	132 ± 3.67	1164.65 ± 1205.20
F × W 151	64 ± 0.00	131 ± 5.00	43.05 ± 0.00
F × W 153	64 ± 0.00	133 ± 4.06	1504.79 ± 1281.49
F × W 157	67.75 ± 2.48	139.25 ± 4.38	1241.07 ± 699.89
F × W 158	67 ± 3.00	136.75 ± 2.48	1294.89 ± 983.78
F × W 173	68.5 ± 3.35	138.25 ± 3.26	1469.27 ± 1149.51
F × W 188	65.25 ± 1.29	136.75 ± 2.48	1314.27 ± 1223.91
F × W 191	64 ± 0.00	143.5 ± 1.50	43.05 ± 21.53
F × W 192	62.75 ± 2.16	132 ± 3.67	939.68 ± 777.59
F × W 200	67 ± 3.00	138.25 ± 3.26	1886.91 ± 928.22
F × W 201	67.75 ± 3.89	135.5 ± 0.50	1652.25 ± 1051.62
F × W 205	65 ± 1.00	129 ± 0.00	64.58 ± 0.00
F × W 215	64.75 ± 1.29	131.25 ± 3.89	1140.96 ± 1075.27
F × W 219	63.25 ± 2.58	133.5 ± 4.50	1351.94 ± 1312.77
F × W 249	-	-	-
F × W 263	65.5 ± 1.50	135 ± 0.00	163.60 ± 81.80
F × W 266	65.5 ± 1.50	135 ± 0.00	2152.77 ± 1016.11
F × W 269	64 ± 0.00	135 ± 0.00	846.04 ± 372.43
F × W 282	65.5 ± 1.50	135 ± 0.00	2975.14 ± 598.46
E × F 2	73 ± 0.00	-	-
E × F 9	66 ± 0.00	142 ± 0.00	43.06 ± 0.00
E × F 10	77 ± 0.00	-	-
E × F 11	-	-	-
E × F 54	-	-	-

Table 16: Field characteristics of F × W125 compared to other FLS-resistant lines

Characteristic	**F×W125**	**Mean Across FLS-Resistant Lines**
Days to 90% Flowering	64.00 ± 0.00	65.54 ± 3.06
Days to 90% Maturity	127.5 ± 1.50	135.02 ± 4.59
Yield	43.05 ± 0.00	1240.33 ± 1184.01

FIGURES

Block 1		Block 2	
F×W	E×F	F×W	E×F

Figure 9: Depiction of field design each year.

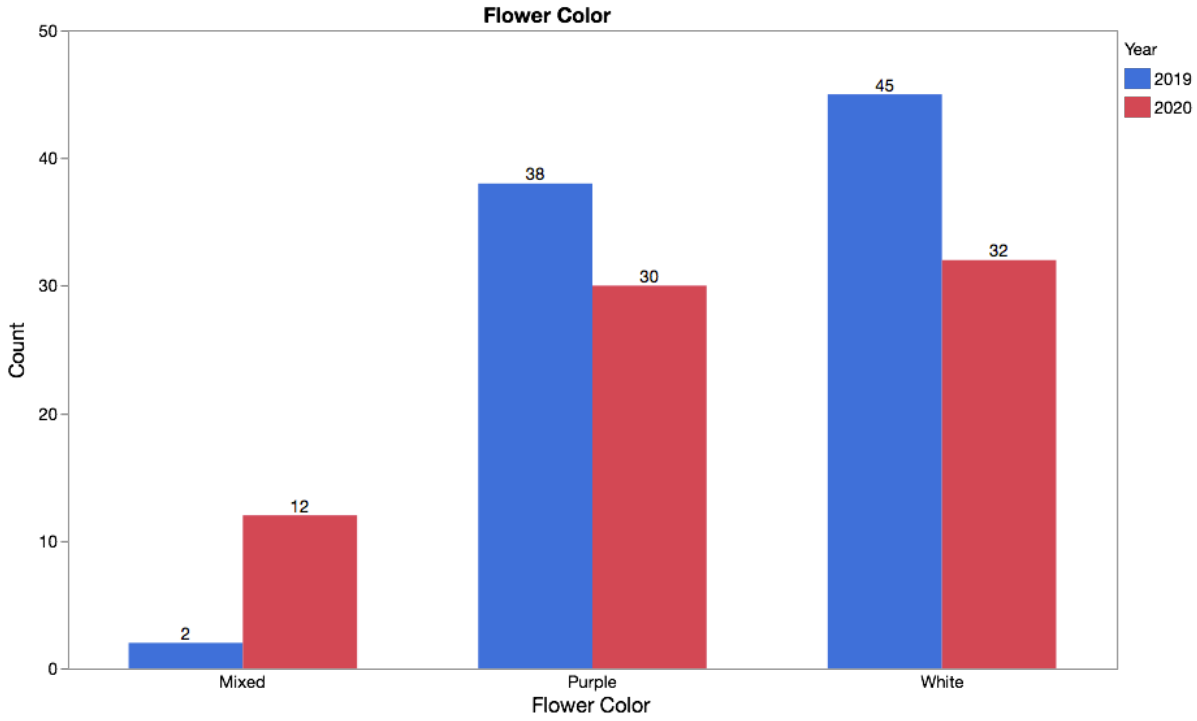


Figure 10: Bar chart comparing flower colors per year.

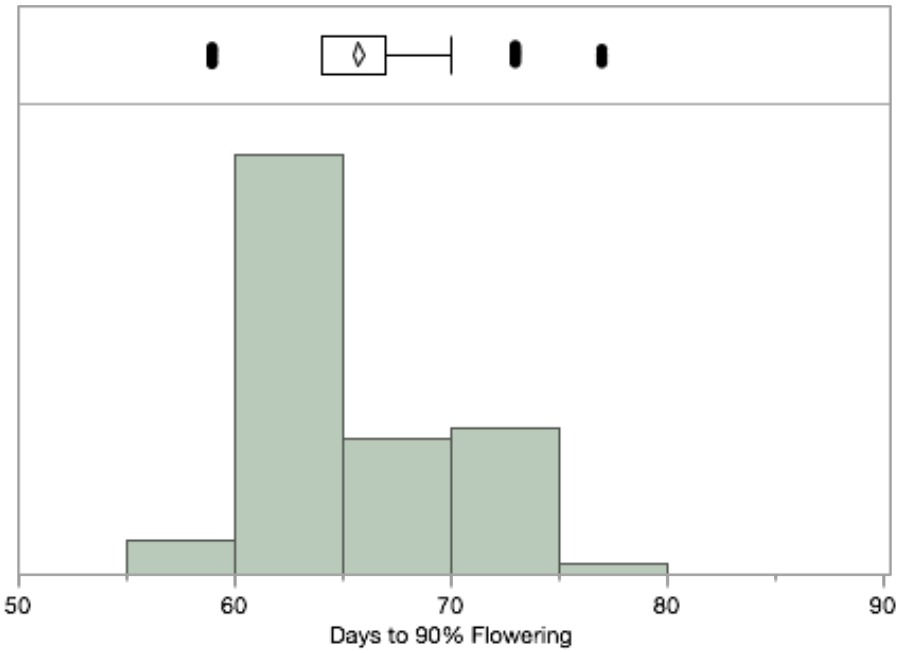


Figure 11: Distribution of the F × W population's days to 90% flowering

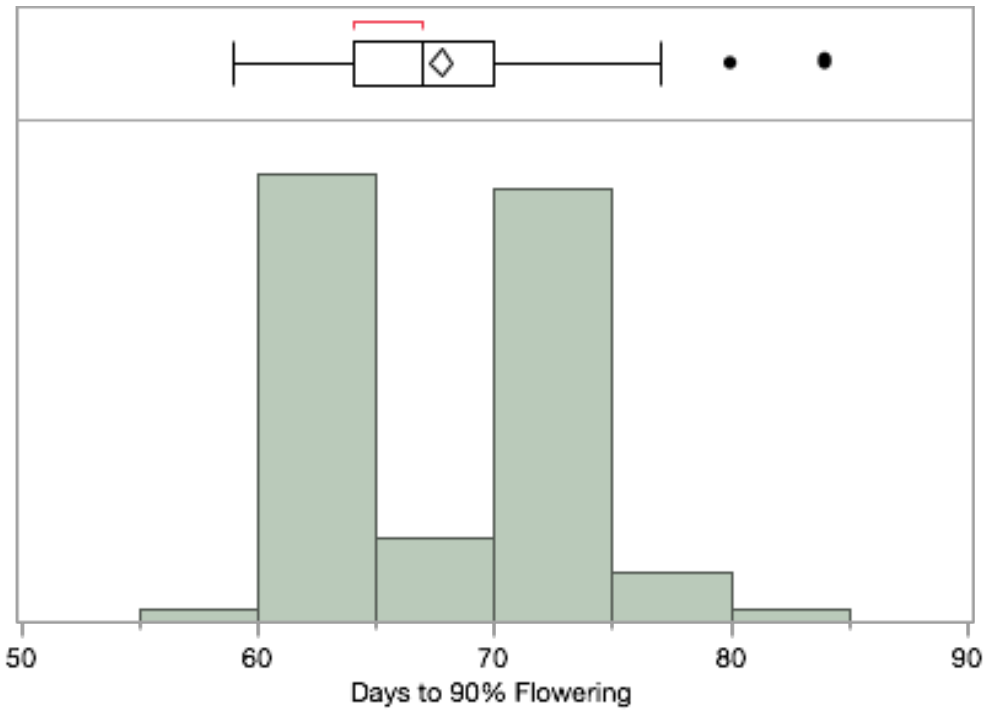


Figure 12: Distribution of the E × F population's days to 90% flowering

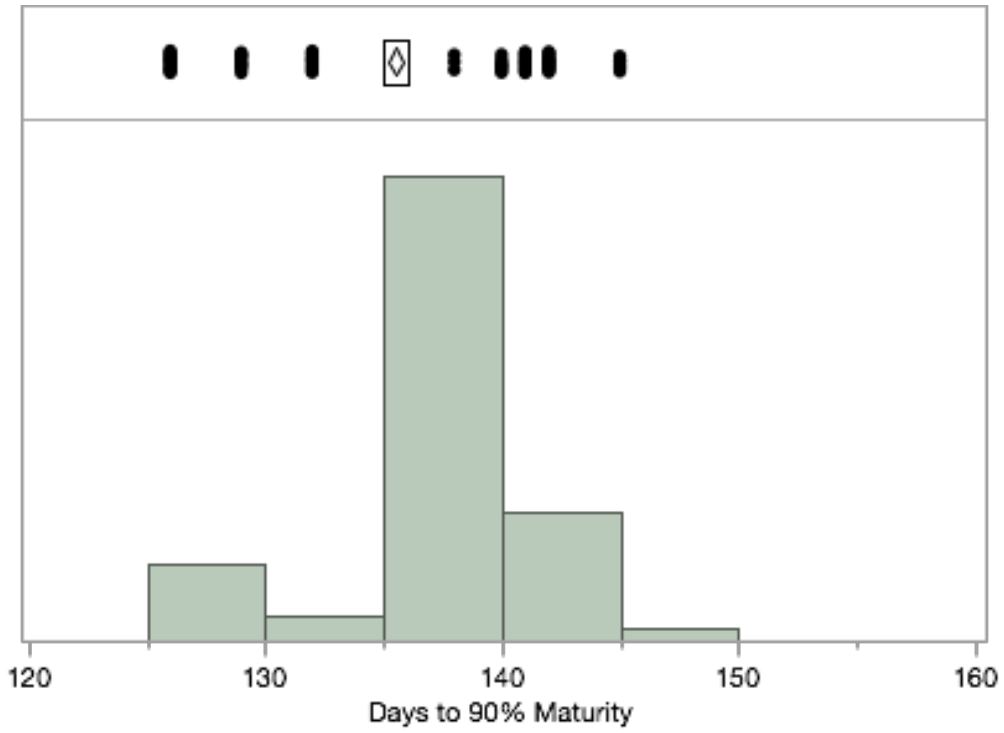


Figure 13: Distribution of the F × W population's days to 90% maturity

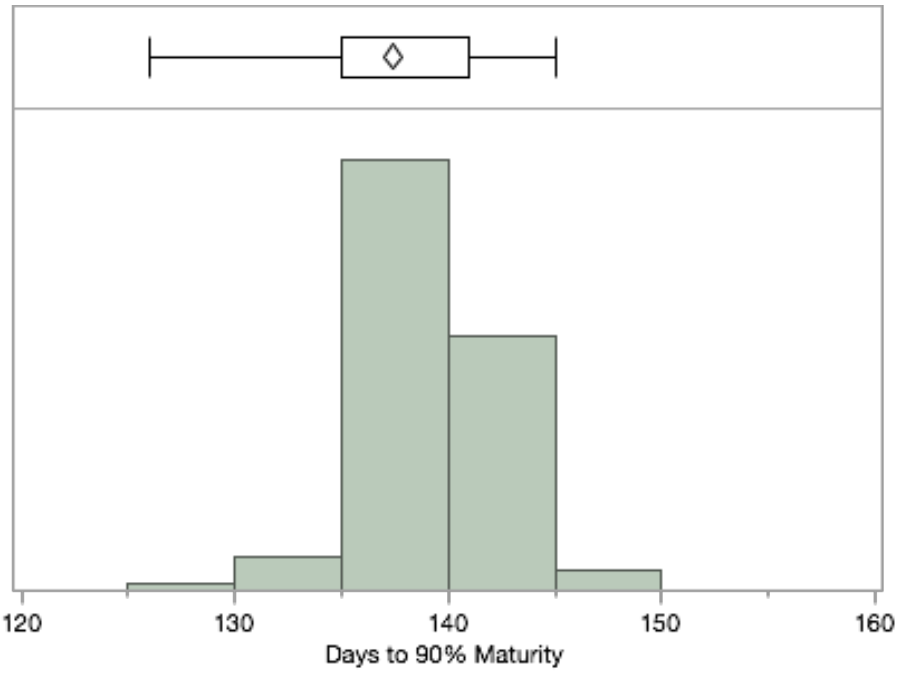


Figure 14: Distribution of the E × F population's days to 90% maturity

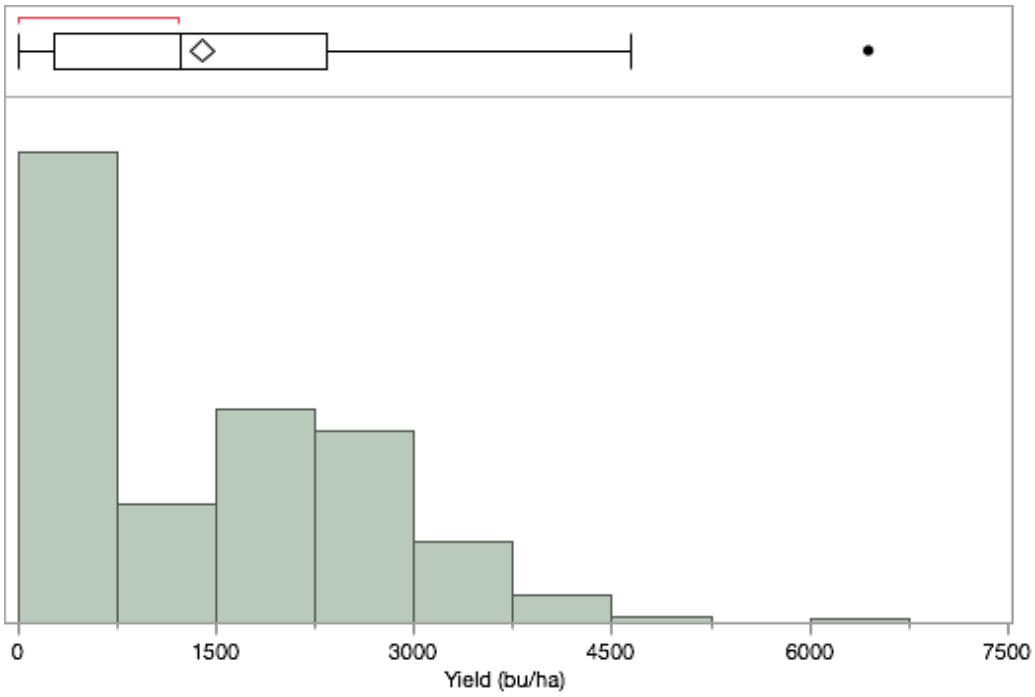


Figure 15: Distribution of the F × W population's yield

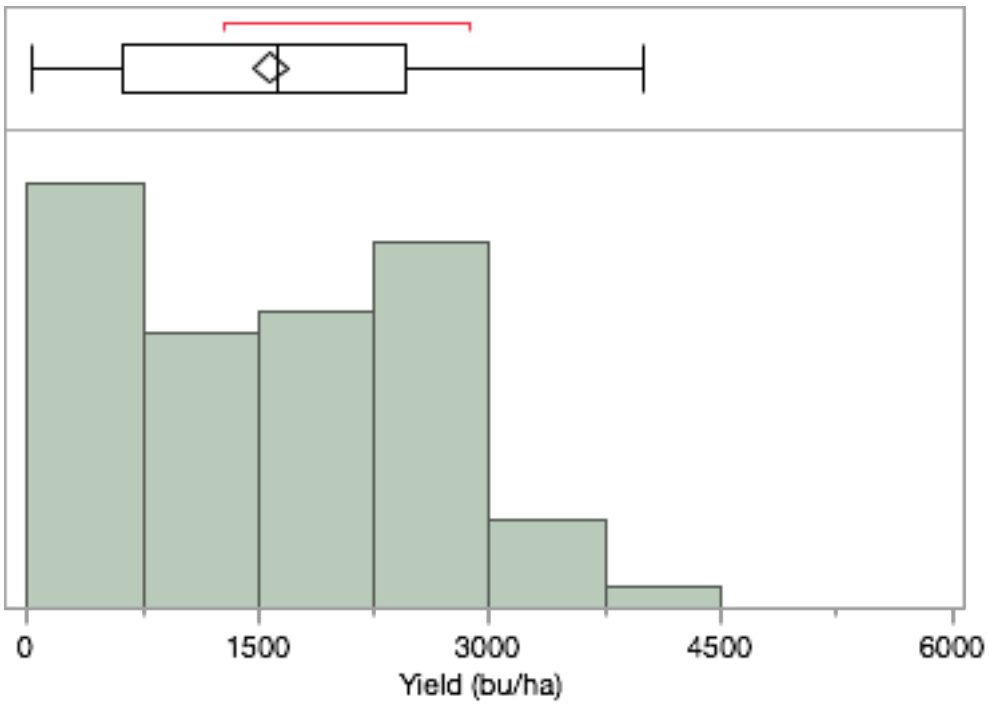


Figure 16: Distribution of the E × F population's yield

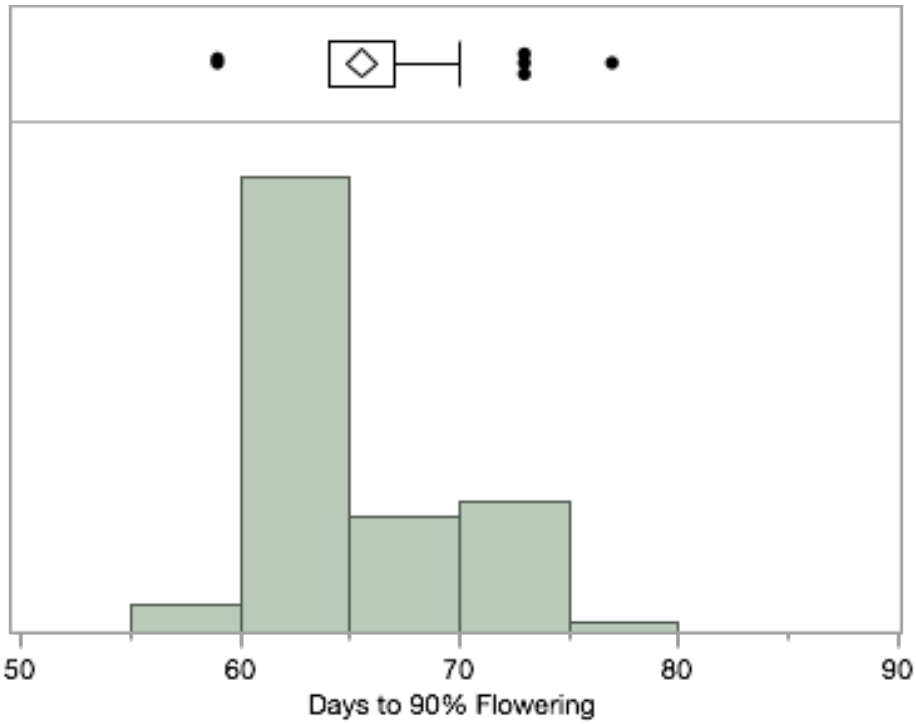


Figure 17: Distribution of FLS-resistant lines' days to 90% flowering

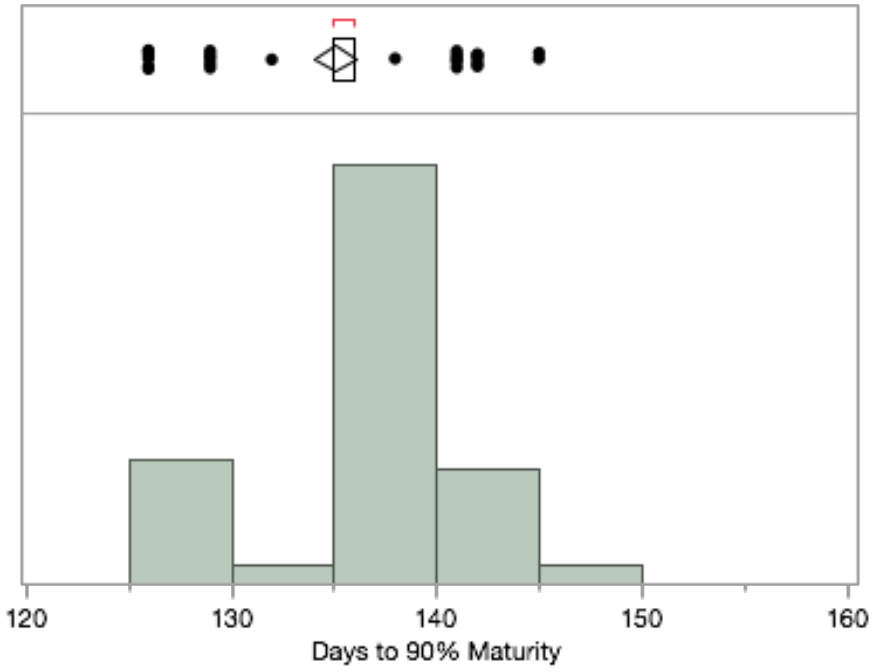


Figure 18: Distribution of FLS-resistant lines' days to 90% maturity

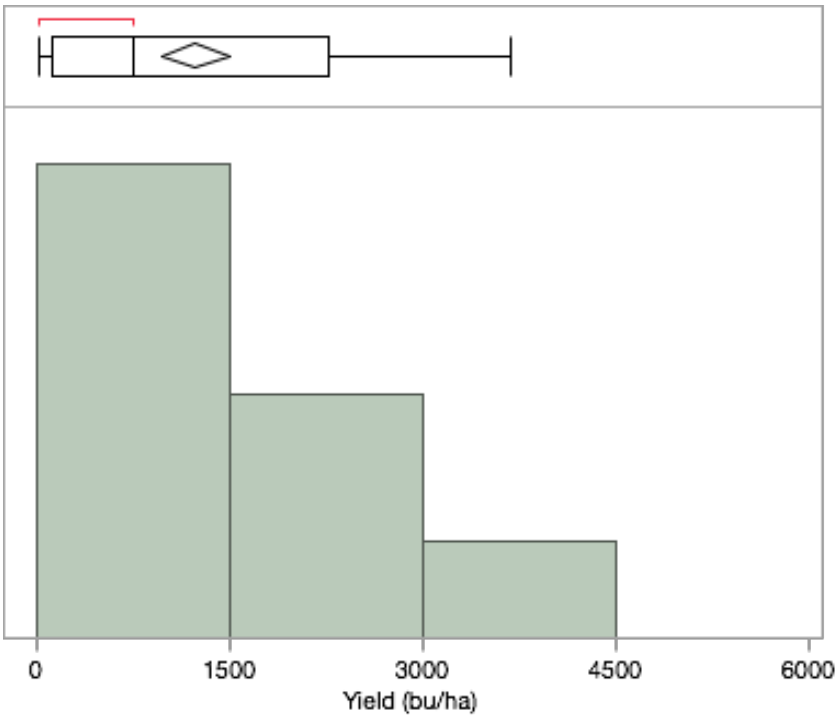


Figure 19: Distribution of FLS-resistant lines' yield

REFERENCES

Allen, T.W., Bradley, C.A., Sisson, A.J., Byamukama, E., Chilvers, M.I., Coker, C.M., Collins, A.A., Damicone, J.P., Dorrance, A.E., Dufault, N.S., Esker, P.D., Faske, T.R., Geisler, L.J., Grybauskas, A.P., Hershman, D.E., Hollier, C.A., Isakeit, T., Jardine, D.J., Kelly, H.M., Kemerait, R.C., Kleczewski, N.M., Koenning, S.R., Kurle, J.E., Malvick, D.K., Markell, S.G., Mehl, H.L., Mueller, D.S., Mueller, J.D., Mulrooney, R.P., Nelson, B.D., Newman, M.A., Osborne, L., Overstreet, C., Padgett, G.B., Phipps, P.M., Price, P.P., Sikora, E.J., Smith, D.L., Spurlock, T.N., Tande, C.A., Tenuta, A.U., Wise, K.A., & Wrather, J.A. (2017). Soybean yield loss estimates due to diseases in the United States and Ontario, Canada, from 2010 to 2014. *Plant Health Progress 18*(1): 19-27. https://doi.org/10.1094/PHP-RS-16-0066

Alves, B.J.R., Boddey, R.M., & Urquiaga S. (2003). The success of BNF in soybean in Brazil. *Plant and Soil 252*, 1-9. https://doi.org/10.1023/A:1024191913296

Atibalentja, N., Jakstys, B.P., & Noel, G.R. (2014). Life cycle, ultrastructure, and host specificity of the North American isolate of *Pasteuria* that parasitizes the soybean cyst nematode, *Heterodera glycines*. *Journal of Nematology 36*(2): 171-180.

Arai, Y., Makoto, H., Nishimura, M. (2008). Proteomic analysis of highly purified peroxisomes from etiolated soybean cotyledons. *Plant and Cell Physiology, 49*(4): 526-539. https://doi.org/10.1093/pcp/pcn027

Aref, S., & Pike, R. (1998). Midwest farmers' perceptions of crop pest infestation. *Agronomy Journal 90*(6): 819-825. https://doi.org/10.2134/agronj1998.00021962009000060013x

Ash, M., Livezey, J., & Dohlman, E. (2006). Soybean Backgrounder. *Electronic Outlook Report from the Economic Research Service 1.* Retrieved from https://www.ers.usda.gov/webdocs/outlooks/38482/29495_ocs200601.pdf?v=9984.5

Bernard, R. L. (1971). Two major genes for time of flowering and maturity in soybeans. *Crop Science 11*(2):242-244. https://doi.org/10.2135/cropsci1971.0011183X001100020022x

Bernard, R.L. (1972). Two genes affecting stem termination in soybeans. *Crop Science 12*(2): 235-239. https://doi.org/10.2135/cropsci1972.0011183X001200020028x

Boerma, H.R., & Specht, J.E. (2004). *Soybeans: Improvement Production, and Uses, Third Edition.* Madison, WI: American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America.

Britannica, The Editors of Encyclopaedia. (2019). *Soybean. Encyclopedia Britannica.* Retrieved from https://www.britannica.com/plant/soybean

Broman, K.W. (2005). The genomes of recombinant inbred lines. *Genetics 169*(2): 1133-1146. https://doi.org/10.1534/genetics.104.035212

Broman, K.W., Sen, Ś. (2009). *A Guide to QTL Mapping with R/qtl*. New York City, NY: Springer-Verlag Publishing.

Broman, K.W., Wu, H., Śaunak, S., Churchill, G.A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics 19*(7): 889-890. https://doi.org/10.1093/bioinformatics/btg112

Byamukama, E., Febina, M., Connie, T., Strunk, C. (2019). *Frogeye leaf spot of soybean*. (Bulletin P-00121) SDSU Extension, South Dakota State University.

Cahal, G.S., Gostal, S.S. (2002). *Principles and procedures of plant breeding: biotechnological and conventional approaches.* Boca Raton, FL: CRC Press.

Carrls, L.M., Little, C.R., & Stiles, C.M. (2012). Introduction to fungi. *The Plant Health Instructor*.

Chang, H.X., Lipka, A.E., Domier, L.L., Hartman, G.L. (2016). Characterization of disease resistance loci in the USDA soybean germplasm collection using genome-wide association studies. *Phytopathology 106*, 1139-1151. https://doi.org/10.1094/PHYTO-01-16-0042-FI

Concibido, V.C., La Vallee, B., Mclaird, P., Pineda, N., Meyer, J., Hummel, L., Yang, J., Wu, K., & Delannay, X. (2003). Introgression of a quantitative trait locus for yield from *Glycine soja* into commercial soybean cultivars. *Theoretical Applied Genetics 106*, 575-582. https://doi.org/10.1007/s00122-002-1071-5

Cook, D.E., Lee, T.G., Guo, X., Melito, S., Wang, K., Bayless, A.M., Wang, J., Huges, T., Willis, D., Clemente, T.E., Diers, B.W., Jiang, J., Hudson, M.E., & Bent, A.F. (2012). Copy number variation of multiple genes at *Rhg1* mediates nematode resistance in soybean. *Science 388*(6111): 1206-1209. https://doi.org/10.1126/science.1228746

Cooper, J. (2007) *Fungal plant pathogens and symptomology*. Retrieved from https://s3.wp.wsu.edu/uploads/sites/2054/2014/04/FungalPlantPathogens_002.pdf

Destro, D., Carpentieri-Pipolo, V., Kiihl, R.A. de S., & Almeida, L.A. de. (2001). Photoperiodism and genetic control of the long juvenile period in soybean: a review. *Crop Breeding and Applied Biotechnology 1*(1): 72-92. http://dx.doi.org/10.13082/1984-7033.v01n01a10

Dhaubhadel, S., Kuflu, K., Romero, M.C., Gijzen, M. (2005). A soybean seed protein with carboxylate-binding activity. *Journal of Experimental Botany 56*(419): 2335-2344. https://doi.org/10.1093/jxb/eri226

Elhady, A., Hallmann, J., & Huer, H. (2020) Symbiosis of soybean with nitrogen fixing bacteria affected by root lesion nematodes in a density-dependent manner. *Scientific Reports 10*, 1619. https://doi.org/10.1038/s41598-020-58546-x

Food and Agriculture Organization of the United Nations (2019). *Soybean production in 2018, crops/world regions/production quantity.* Retrieved from http://www.fao.org/faostat/en/#data/QC/

Gantt, J., Key, J. (1985). Coordinate expression of ribosomal protein mRNAs following auxin treatment of soybean hypocotyls. *Journal of Biological Chemistry 260*(10):6175-6181.

Gordon, S.G., Martin, S.K.S., Dorrance, A.E. (2006). Rps8 Maps to a resistance gene rich region on soybean molecular linkage group F. *Crop Science 46*, 168-173. https://doi.org/10.2135/cropsci2004.04-0024

Grant, D., Nelson, R.T., Cannon, S.B., Shoemaker, R.C. (2010). SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Research 38*(1): D843-D846. https://doi.org/10.1093/nar/gkp798

Gururani, M.A., Venkatesh, J., Upadhyaya, C.P., Nookaraju, A., Pandey, S.K., Park, S.W. (2012). Plant disease resistance genes: current status and future directions. *Physiological and Molecular Plant Pathology 78*(2012): 51-65. https://doi.org/10.1016/j.pmpp.2012.01.002

Hartman, G.L., Sinclair, & J.B., Rupe, J.C. (ed.) (1999). *Compendium of soybean diseases. 4th ed.* St. Paul, MN: APS Press.

Hill, J.H., & Whitham, S.A. (2014). Chapter seven – control of virus diseases in soybeans. *Advances in Virus Research* (pp. 355-390). Academic Press. https://doi.org/10.1016/B978-0-12-801246-8.00007-X

Hoskins, A.J. (2011). *Genetic mapping of soybean resistance genes to frogeye leaf spot in five Chinese plant introductions and efficiency of early generation selection for low phytate soybean lines.* (Doctoral dissertation, University of Georgia). Retrieved from https://getd.libs.uga.edu/pdfs/hoskins_aaron_j_201105_phd.pdf

Hymowitz, T., *Soybeans: the success story*. Portland, OR: Timber Press.

Intel, M. (2020). 2020 corn and soybean yield expectations will be inevitably lower in September. Retrived from https://www.fb.org/market-intel/2020-corn-and-soybean-yield-expectations-will-inevitably-lower-in-september#:~:text=The%202019%20soybean%20yield%20was,at%2049.5%20bushels%20per%20acre.

Krause, M.K., & Brand, E.V. (2016). Chapter 9- Scallop genetics and genomics. In Shumway, S.E., & Parsons, G.J. (Eds.), *Developments in Aquaculture and Fisheries Science* (pp. 371-424). Elsevier. https://doi.org/10.1016/B978-0-444-62710-0.00009-2

Laur, J., Porter, P., & Oplinger, E. (1997). The corn and soybean rotation. *Field Crops 27*, 426.

Lee, G.A., Crawford, G.W., Li, L., Sasaki, Y., & Chen, X. (2011). Archaeological soybean *(Glycine max)* in east Asia: Does size matter? *PLoS One 6*(11): e26720. https://doi.org/10.1371/journal.pone.0026720

Lehman, S.G., (1928). Frogeye leaf spot of soybean caused by Cercospora diazu Miara. *Journal of Agricultural Research 36*(9):811-833.

Lenvik, T.R., Key, J.L., Gantt, J.S. (1994). Ribosomal protein S11 genes from Arabidopsis and soybean. *Plant Physiology 105*(3):1027-1028. https://doi.org/10.1104/pp.105.3.1027

Lightfoot, D.A., Nijiti, V.N., Gibson, P.T., Kassem, M.A., Iqbal, J.M., Meksem, K. (2005). Registration of the Essex x Forrest recombinant inbred line mapping population. *Crop Science 45*, 1678-1681. https://doi.org/10.2135/cropsci2004.0279

Lightfoot, D.A. (2008). Soybean genomics: developments through the use of cultivar "Forrest." *International Journal of Plant Genomics 2008*. https://doi.org/10.1155/2008/793158

Ling, M.M., & Robinson, B.H. (1997). Approaches to DNA mutagenesis: an overview. *Analytical Biochemistry 254*(2): 157-178. https://doi.org/10.1006/abio.1997.2428

Lin, B., & Kelly, H. (2018). Frogeye leaf spot of soybean. *The Plant Health Instructor*. https://doi.org/10.1094/PHI-I-2018-1018-01

Malvick, D. (2018). *Frogeye leaf spot on soybean*. Retrieved from https://extension.umn.edu/pest-management/frogeye-leaf-spot-soybean#:~:text=Symptoms%20of%20frogeye%20leaf%20spot,reddish%2Dbrown%20or%20purple%20ring.

Mammadov, J., Aggarwal, R., Buyyarapu, R., & Kumpatla, S. (2012). SNP markers and their impact on plant breeding. *International Journal of Plant Genomics 2012*, 1-11. https://doi.org/10.1155/2012/728398

Mengistu, A., Kurtzweil, N.C., & Grau, C.R. First report of frogeye leaf spot (*Cercospora sojina)* in Wisconsin. *Plant Disease 86*(11). https://doi.org/10.1094/PDIS.2002.86.11.1272B

Mian, R., Missaoui, A.M., & Walker, D.R. (2007). Frogeye leaf spot of soybean: a review and proposed race designations for isolates of hara. *Crop Science 48*(1): 14-24. https://doi.org/10.2135/cropsci2007.08.0432

Miransari, M. (2016). Soybeans, stress, and plant growth- promoting rhizobacteria. *Environmental Stresses in Soybean Production* (pp. 177-203). Elsevier Inc. https://doi.org/10.1016/B978-0-12-801535-3.00008-5

Mourtzinis, S., & Conley, S.P. (2017). Delineating soybean maturity groups across the United States. *Agronomy Journal 109*(4): 1397-1403. https://doi.org/10.2134/agronj2016.10.0581

National Human Genome Research Institute. (2020). *Restriction fragment length polymorphism (RFLP).* Retrieved from https://www.genome.gov/genetics-glossary/Restriction-Fragment-Length-Polymorphism

Negrutiu, I., Hinnisdaels, S., Mouras, A., Gill, B.S., Gharti-Chhetri, G.B., Davey, M.R., Gleba, Y.Y., Sidorov, V., & Jacobs, M. (1989). Somatic versus sexual hybridization: features, facts, and future. *Acta botanica neerlandica 38*(3): 253-272.

Niblack, T.L. (2007). Soybean cyst nematode management reconsidered. *Plant Disease 89*(10). https://doi.org/10.1094/PD-89-1020

Pham, A.T., Harris, D.K., Buck, J., Hoskins, A., Serrano, J., Abdel-Haleem, H., Cregan, P., Song, Q., Boerma, H.R., Li, Z. (2015). Fine mapping and characterization of candidate genes that control resistance to *Cercospora* sojina K. Hara in two soybean germplasm accessions. PLoS *ONE 10*(5): e0126753. https://doi.org/10.1371/journal.pone.0126753

Philips, D.V., Boerma, H.R. (1981). Cercospora sojina race 5: a threat to soybeans in the southeastern United States. *American Phytopathological Society 71*, 334-336.

Purcell, L.C., Salmeron, M., & Ashlock, L. (2014). Soybean Growth and Development. *Arkansas Soybean Production Handbook* (pp. 1-8). Little Rock, AR: University of Arkansas Division of Agriculture Research & Extension.

Reeve, J.D. (2016). Expected Value, Variance, and Samples. *Biostatistics: Data and Models* (pp.181-204). https://biostatisticsjdreeve.files.wordpress.com/2016/07/chapter-7-expected-value.pdf

Salas, B.S., Secor, G.A., Taylor, R.J., Gudmestad, N.C. (2007). Assessment of resistance of tubers of potato cultivars to *Phytophthora erythroseptica* and *Pythium ultimum*. *Plant Disease 87*(1):91-97. https://doi.org/10.1094/PDIS.2003.87.1.91

Sharma, H., Lightfoot D.A. (2017). Quantitative trait loci underlying partial resistance to *Cercospora sojina* race 2 detected in soybean seedlings in greenhouse assays. *Atlas Journal of Biology 3*, 175-182. https://doi.org/10.5147/ajb.v3i1.29

Sinclair, J.B. (1982). Introduction Pages in: Compendium of Soybean Diseases 2nd Edition (1-2). St. Paul, MN: American Phytopathological Society Press.

Singh, G. (2010). *The Soybean: Botany, Production and Uses.* United Kingdom: CABI.

Singh, N., Choudhury, D.R., Singh, A.K., Kumar, S., Srinivasan, K., Tyagi, R.K., Singh, N.K., Singh, R. (2013). Comparison of SSR and SNP markers in estimation of genetic diversity and population structure of Indian rice varieties. *PLOS ONE 8*(12): e84136. https://doi.org/10.1371/journal.pone.0084136

Staton, M. (2012). *Managing soil pH for optimal soybean production.* Retrieved from https://www.canr.msu.edu/news/managing_soil_ph_for_optimal_soybean_production#:~:text=Soybeans%20will%20generally%20perform%20well,nematode%20(SCN)%20population%20growth.

The UniProt Consortium (2020). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research 47*(D1): D506-D515. https://doi.org/10.1093/nar/gky1049

Voora, V., Larrea, C., & Bermudez, S. (2020). *Global Market Report: Soybeans*. Retrieved from https://www.iisd.org/system/files/2020-10/ssi-global-market-report-soybean.pdf

Wantanabe, S., Hideshima R., Xia Z., Tsubokura Y., Sato S., Nakamoto Y., Yamanaka N., Takahashi R., Ishimoto M., Anai T., Tabata S., & Harada K. (2009). Map-based cloning of the gene associated with the soybean maturity locus E3. *Genetics 182*(4): 1251-1262. https://doi.org/10.1534/genetics.108.098772

Wantanabe, S., Harada K., & Abe J. (2012). Genetic and molecular bases of photoperiod responses of flowering in soybean. *Breeding Science 61*(5): 531-543. https://doi.org/10.1270/jsbbs.61.531

Wrather, J.A., Anderson, T.R., Arsyad, D.M., Tan, Y., Ploper, L.D., Porta-Puglia, A., Ram, H.H., Yorinori, J.T. (2001). Soybean disease loss estimates for the top ten soybean-producing countries in 1998. *Canadian Journal of Plant Pathology 23*, 115-121. https://doi.org/10.1080/07060660109506918

Wu, X., Vuong, T.D., Leroy, J.A., Shannon, J.G., Sleper, D.A., Nguyen, H.T. (2011). Selection of a core set of RILs from Forrest x Williams 82 to develop a framework map in soybean. *Theoretical and Applied Genetics 122*(6):1179-1187. https://doi.org/10.1007/s00122-010-1522-3

Yang, W., Zou, J., Wang, J., Li, N., Luo, X., Jiang, X., Li, S. (2020). Wide crossing diversify mitogenomes of rice. *BMC Plant Biology 20*(159). https://doi.org/10.1186/s12870-020-02380-w

Yin, X., Stam, P., Kropff, M.J., Schapendonk, H.C.M. (2003). Crop modeling, QTL mapping, and their complementary role in plant breeding. *Agronomy Journal 95*(1): 90-98. https://doi.org/10.2134/agronj2003.9000a

Yong, H. (1767). *The Universal Magazine of Knowledge and Pleasure, 40*(279), 266.

Yousef, G.G., Juvik, J.A. (2001). Comparison of phenotypic and marker-assisted selection for quantitative traits in sweet corn. *Crop Breeding, Genetics & Cytology 41*(3): 645-655. https://doi.org/10.2135/cropsci2001.413645x

Zhang, G. (2012). *Cercospora sojina: over-winter survival and fungicide resistance*. (Doctoral dissertation, University of Illinois at Urbana-Champaign). Retrieved from https://www.ideals.illinois.edu/bitstream/handle/2142/31005/Zhang_Guirong.pdf?sequence=1&isAllowed=y

Zhang, G., Allen, T.W., Bond, J.P., Fakhoury, A.M., Dorrance, A.E., Weber, L., Faske, T.R., Geisler, L.J., Hershman, D.E., Kennedy, B.S., Neves, D.L., Hollier, C.A., Kelly, H.M., Newman, M.A., Kleczewski, N.M., Koenning, S.R., Thiessen, L.D., Mehl, H.L., Zhou, T., Meyer, M.D., Mueller, D.S., Kandel, Y.R., Price, P.P. III, Rupe, J.C., Sikora, E.J., Standish, J.R., Tomaso-Petersoon, M., Wise, K.A., & Bradley, C.A.(2018). Widespread Occurrence of quinone outside inhibitor fungicide-resistant isolates of *Cercospora sojina,* causal agent of frogeye leaf spot of soybean, in the United States. *Plant Health Progress 19*(4): 295-302. https://doi.org/10.1094/PHP-04-18-0016-

VITA

Graduate School
Southern Illinois University

Kelsey R. Smith

2015ksmith@gmail.com

Southern Illinois University Carbondale
Bachelor of Science, Crop, Soil, and Environmental Management, May 2019

Thesis Paper Title:
Identifying Frogeye Leaf Spot Resistance in Two Elite Soybean Populations and Analysis of Agronomic Traits in Resistant Lines

Major Professor: Dr. Stella Kantartzi

Publications:

McAllister, K.R., Lee, Y.-C., & Kantartzi, S.K. (2021). QTL mapping for resistance to *Cercospora sojina* in 'Essex' × 'Forrest' soybean (*Glycine max* L.) lines. *Journal of Plant Breeding and Crop Science 13*(1): 14-22. https://doi.org/10.5897/JPBCS2020.0938