BREEDING FOR INCREASED PROTEIN CONCENTRATION, SUDDEN DEATH SYNDROME RESISTANCE, AND SOYBEAN CYST NEMATODE RESISTANCE IN SOYBEAN

BY

LILLIAN FRANCES BRZOSTOWSKI

DISSERTATION

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

Urbana, Illinois

Doctoral Committee:

Professor Brian Diers, Chair
Professor Glen Hartman
Professor Fred Kolb
Professor Randy Nelson

ABSTRACT

Soybean (*Glycine max* (L.) Merr.) is one of the most important agronomic crops in the USA and worldwide with dynamic uses in food, industry, and feed. Development of improved soybean cultivars is critical to provide the resources necessary for a growing world population. Approximately 83 million acres/33.5 million hectares of soybean were harvested in 2016 in the USA, and Illinois is one of the top soybean producing states. For the past several decades, soybean breeders have sought to protect and improve the economic value of soybean through genetic improvement of seed composition and disease resistance traits. In order for a gene to be effectively incorporated into a breeding program, it must maintain its desired effect across many genetic backgrounds without a negative effect on agronomic traits such as yield. The objective of this dissertation was to identify genetic regions that can be used in breeding programs to successfully increase protein concentration, sudden death syndrome (SDS) resistance, and soybean cyst nematode (SCN) resistance.

ACKNOWLEDGEMENTS

I would like to extend heartfelt thanks to my advisor, Dr. Brian Diers, for his guidance during my PhD. I would also like to thank my committee members Drs. Glen Hartman, Fred Kolb, and Randy Nelson for providing me feedback and challenging me. Additionally, much of my work could not have been completed without the help of collaborators including Roger Bowen, Cathy Schmidt, Jason Bond, Silvia Cianzio, Dechun Wang, Alison Colgrove, and Andrew Scaboo. I am especially grateful to the United Soybean Board with special acknowledgement to the soybean farmers across the United States and also to Dr. Robert Fraley and his family for providing funding for my fellowships. Sincere gratitude also goes to Troy Cary, Sarah Schultz, John Meharry, and Jake Tammen along with past and present Diers lab undergraduate students, graduate students, and visiting scientists. I also thank my colleagues at UIUC that became good friends: Elizabeth Hawkins, Jessica Bubert, Nathan Schmitz, and Russell Ward.

Finally, I would not be where I am today without the constant support from my family namely Mom, Dad, Mumzy, Pop Pop, Jane, and Mike. Most of all, I would like to thank my wonderful husband Rob Ziegler and my two little bubbas, Stubby and Joe. I love you so much, and I would not be where I am today without you.

TABLE OF CONTENTS

CHAPTER 1: LITERATURE REVIEW OF BREEDING FOR INCREASED PROTEIN CONCENTRATION, SUDDEN DEATH SYNDROME RESISTANCE, AND SOYBEAN CYST NEMATODE RESISTANCE IN SOYBEAN

Soybean

Approximately 3,000 to 5,000 years ago, cultivated soybean was domesticated in China (Carter et al., 2004). The first recorded production in North America was in 1765 when a few landraces were introduced to North America (Hymowitz and Harlan, 1983). Soybean was mainly a forage crop until the mid 20th century. During this time, farmers grew plant introductions (PI) from East Asia or selections from these lines (Probst and Judd, 1973). In the 1930s, breeding programs developed cultivars derived from crosses between PIs, and the shift from soybean as a forage to soybean as a grain crop began as there was an increase in demand for soybean oil (Hartwig, 1973). As soybean became a more widely grown crop in the USA, commercial breeding programs were established. With 80 years of selective breeding, the genetic base of North American cultivars is narrow (Gizlice et al., 1994; Sneller, 1994; Hyten et al., 2006). Almost all soybean varieties released between 1947 and 1988 trace their ancestry back to just 80 accessions (Gizlice et al., 1994).

Soybean (*Glycine max* (L.) Merr.) is an important agronomic crop in the US and around the world. It is a dynamic crop with many food, industrial, and feed uses. Processed soybeans are the largest source of protein feed and the second largest source of vegetable oil in the world. Soybean accounts for 90 percent of total oilseed production in the USA (USDA-ERS, 2013).

In 2016, producers harvested over 4.3 billion bushels/117 million metric tons of soybean from over 82.7 million acres/33.5 million hectares in the USA. Yields averaged 52.1 bushels per acre/3501 kg ha^{-1}, and prices were approximately ten dollars a bushel (USDA-NASS, 2017). Over 44 percent of the US soybean and soybean products were exported. China is the top buyer of US whole soybeans and soybean oil, and Mexico is the top buyer of US soybean meal (USDA-NASS, 2017).

In Illinois, soybean is the second most widely grown crop behind corn. In 2016, over 547 million bushels/14.9 million metric tons from 9.77 million acres/3.96 million hectares were harvested in the state (USDA-NASS, 2017). Average yield across Illinois in 2016 was 56 bushels

per acre/3762 kg ha^{-1}. With higher prices than corn and high domestic and international demand, soybean will continue to be major crop for producers (USDA-ERS, 2013).

Soybean Protein Concentration

Part of the economic value of soybean can be attributed to the protein found in its seed. A soybean seed is on average 350 g kg^{-1} protein (130 g kg^{-1} moisture) and contains most of the amino acids that are needed to meet the dietary requirements of swine and poultry (Liu, 1997; Cromwell, 2012). Soybean meal (SBM), which is made out of crushed seed after oil is extracted from it, has the highest level of crude protein compared to other vegetable protein sources and accounts for approximately 69% of protein sources in animal feed worldwide (Wilson, 2008). Over 75% of SBM produced in the United States is fed to poultry or swine; however, SBM is also used as a protein source in the dairy, cattle, pet, and aquaculture industry (Cromwell, 2012; ISA, 2013). With SBM's popularity in the feed industry, it is important for breeders to release cultivars that are not only high yielding for producers but also have a high protein concentration. This will allow the needs of the largest soybean customer, livestock producers, to be met (USDA-ERS, 2012).

Soybean producers currently sell their product to the elevator by weight. They are not paid for component levels; however, research demonstrates market price moves with quality (ISA, 2013). On the whole, prices are lowered with low quality soybean seeds that are in poor condition with low seed composition levels. When this is considered, it becomes advantageous to farmers to grow soybean seeds with a good balance of protein, oil, and amino acids.

Elevators also sell soybean to crush plants by the bushel. The price of SBM is reduced if it does not meet a 48% protein level, so the drive for high protein soybeans come from the processors. The combined value of protein, oil, and hulls is called its Estimated Process Value (EPV) or "crush value." High protein soybeans have a higher EPV and are worth more to the processor. Brazilian producers tend to grow soybeans with higher EPVs, which give them an advantage in the world SBM market. It is important to grow and develop high protein, high yielding cultivars to keep soybean production in the USA economically competitive (ISA, 2013).

Storage proteins, which account for a large portion of protein in soybean, are either water soluble albumins or saline soluble globulins (Murphy, 2008). Most soybean proteins are

globulins which are either 7S vicilin-type and 11S leguminin-type (Clarke and Wiseman, 2000). The 11S faction is represented by glycinin, and the 7S faction is represented by β-conglycinin. (Yaklich et al., 1999).

Relationship between protein and other seed components

Soybean seeds on average are 350 g kg^{-1} protein, 175 g kg^{-1} oil, 260 g kg^{-1} carbohydrates, 45 g kg^{-1} crude fiber, and 45 g kg^{-1} ash (130 g kg^{-1} moisture basis)(Hymowitz, 1972; Wilson, 2004). Complex relationships exist between these components and with seed yield. An increase in protein often leads to a decrease in oil, carbohydrates, and yield.

The inverse relationship between protein and oil concentration has been well established (Hartwig and Kilen, 1990; Wilson, 2004). Wilcox and Shibles (2001) crossed parents with low (345 g kg^{-1} on a 130 g kg^{-1} moisture basis)) and high (413 g kg^{-1}) protein concentration in a study to determine the relationship between protein concentration and oil along with other seed traits including yield, sugar concentration, and sulfur concentration. A highly negative correlation (r=-0.88) between protein and oil concentration was observed. In a large data set from the US Soybean Uniform test representing 10 maturity groups, Pearson correlation coefficients for percentage oil versus protein ranged from -0.1042 to -0.6289 with an average of -0.4273 (Piper and Boote, 1999). Possible reasons for these strong negative correlations include pleiotrophy or a tight repulsion linkage between protein and oil quantitative trait loci (QTL). Recker et al. (2014) conducted 26 generations of random mating in two populations and observed significantly large negative genotypic correlations (>|0.5|) despite theoretically small linkage disequilibrium.

The inverse relationship between protein and oil has been expressed in terms of energetic cost. It had been suggested the energy cost to synthesize two units of protein was the same as the cost of one oil unit (Hanson et al., 1961). Usually, a 1.5 to 1.7 ratio is observed suggesting a genotype dependent ratio or that more energy is required to synthesize protein than previously thought (Hanson et al., 1961, Leffel and Rhodes, 1993; Chung et al., 2003).

An inverse relationship between seed yield and protein concentration has been observed frequently (Burton, 1987; Wilcox and Guodong, 1994). Wilcox and Cavins (1995) observed this trend in summarizing three breeding studies where the inverse relationship between protein concentration and yield had r values between -0.23 and -0.86. In the previously mentioned

Wilcox and Shibles (2001) study, there was a 3.9 g kg^{-1} decrease in protein for each 100 kg ha^{-1} yield increase and the correlation was -0.45. Despite the negative correlation between yield and protein being well documented, the reason for the relationship remains unclear. It has been hypothesized the negative correlation is the result of the physiological relationship between N and carbon supply. Both components play a major role in yield and protein concentration. A relationship between carbon and N may exist as they are both affected by N accumulation, partitioning, and whole plant remobilization. Additionally, protein synthesis in soybean requires a large amount of nitrogen in comparison to other legumes and vegetables (Sinclair and deWitt, 1975).

Hanson et al (1961) explained the relationship between protein, oil, and carbohydrates in terms of a seed's energy commitment using a regression model. Protein and oil each use the same limited carbon supply. Based on the regression model, the protein fraction took 0.7863 grams carbon energy to produce 1 gram of protein; the oil fraction took 1.1423 grams carbon energy to produce 1 gram of oil; and, the residual fraction (carbohydrates) took 0.400 grams carbon energy to produce 1 gram of residual. A high yielding, high protein soybean should be possible if nitrogen is not limiting because these factions do not require as much carbon energy as oil.

Later work with the energy investment concept supported an alternative hypothesis where in a high energy environment, the residual fraction is supported first, then the oil component, and then the protein component. A miscalculation in the 1961 formula or a physiological barrier to high protein production was suggested (Shimura and Hanson, 1970). It takes less energy to produce a gram of protein than a gram of oil, but the advantage is negated by a decrease in the residual fraction. This could be a reason protein increases at the expense of yield (Hanson, 1991). Chung et al. (2003) also mentioned the carbon supply limitation as a reason for the inverse relationship between protein and oil. It was also suggested a lack of genetic diversity could explain why more carbohydrates are not metabolized into protein and oil. Additionally, it is possible that there is a limitation to how much photosynthate a plant can produce and move to the seed.

There is also a negative relationship between protein concentration and sugars. Hartwig et al (1997) measured raffinose, stachyose, and sucrose among 20 high protein lines and observed a strong negative relationship between sucrose and protein (r=-0.78). There was also a

negative correlation between protein and raffinose (r=-0.26) A negative correlation was observed between protein and stachyose + raffinose, but it was not significant. Stachyose and raffinose have a detrimental effect on the nutritional qualities of animal feed (Hartwig et al., 1997). Wilcox and Shibles (2001) saw similar results in 43 random breeding lines. Protein increased whereas oil, total carbohydrates, stachyose + raffinose, and sucrose all decreased. A negative decrease in sugars is expected with an increase in protein concentration as protein, oil, and the seed coat account for 70% of seed weight (Leffel and Rhodes, 1993). When protein is increased, oil decreases; however, protein + oil increases. The seed coat remains the same, so the residual seed faction, which includes sugar concentration, must decrease (Hartwig et al., 1997).

Environment and soybean protein concentration

Several studies have focused on environmental conditions that effect protein concentration. The effect of water supply and temperature has been of specific interest to researchers. Rotundo and Westgate (2009) conducted a meta-analysis of environmental effects including drought and temperature stress that affect final seed composition. Generally, they found drought decreased protein concentration. Across all studies reviewed a relative reduction of 16% of total protein concentration was observed when drought occurred. A few studies where protein concentration increased under drought were noted. In these instances, plants were subjected to drought at early reproductive stages. A possible reason for this increase is related to yield components. Drought in the early reproductive stages leads to fewer seeds per plant. As a result, larger seeds develop with a greater source to sink ratio occuring during seed fill (Borras et al., 2004).

Specht et al. (2001) sought to determine the genetics underlying drought tolerance in a field population of $F_{7:11}$ recombinant inbred lines (RILs). The second year of the study saw severe water limitations coupled with hot temperature and winds. Seed protein concentration decreased whereas oil concentration increased with more drought during that year. Carrera et al. (2009) related temperature and drought to protein and oil concentration in 82 soybean multi-environment trials across Argentinian soybean growing areas using multiple linear regression. High temperatures during seed fill combined with ample water led to a small decrease in protein. Within the limited water environments, protein concentration decreased with increasing water

deficits (Carrera et al., 2009). In contrast, Dornbos and Mullen (1992) observed severe drought stress (50% soil saturation with either 33°C or 35°C air temperature) increased protein concentration 4.4% (Dornbos and Mullen, 1992).

Bellaloui and Mengistu (2008) observed the effect of full-season irrigation, reproductive stage irrigation, and non-irrigation on seed composition of the cultivars "Dwight" and "Freedom." Full season irrigation increased protein concentration in Dwight compared to non-irrigation whereas the opposite effect was observed in Freedom. The observed genotype by treatment interaction suggests protein concentration is a complex trait affected by many factors.

Carter et al (1986) observed that the decrease in protein concentration with increasing temperatures (18°C, 22°C, 26°C, 30°C) is not the same for all genotypes. In NC-111, protein concentration progressively decreased from 584 g kg^{-1} at 18°C to 438 g kg^{-1} at 30°C; however, on a per seed basis, protein content remained the same across treatments. In NC-106, protein concentration decreased on a smaller scale from 568 g kg^{-1} at 18°C to 438 g kg^{-1} at 30°C whereas on a per seed basis there was a greater decrease. The content per seed was 139.9 mg $seed^{-1}$ at 18°C and 103.4 mg $seed^{-1}$ at 30°C (Carter et al, 1986). Gibson and Mullen (1996) looked at the effect of mean daily temperatures of 25°C, 27.5°C, 30°C, and 32.5°C on protein concentration in a growth chamber experiment. The temperature treatment only had a minor effect during R1-R5 growth stages, but increasing mean daily temperatures increased protein concentration during R5-R8 (Gibson and Mullen, 1996). Dornbos and Mullen indicated a quadratic model was appropriate to model protein concentration and temperature. As the mean daily temperature increased from 20-27°C, protein concentration decreased whereas as the daily mean temperature increased from 27°C -35°C, protein concentration increased (Dornbos and Mullen, 1992). Conflicting results on temperatures impact on protein concentration support the idea that factors such as timing of stress, genotype, and other environmental conditions are all important in determining final protein concentration.

Germplasm and soybean protein concentration

There is protein concentration diversity in soybean germplasm. In the USDA Soybean Germplasm Collection, 130 g kg^{-1} moisture basis protein concentration is between 276 g kg^{-1} and

504 g kg^{-1}. Out of 14820 accessions in the collection with protein phenotypic data, 13,697 have a concentration greater than 348 g kg^{-1} (USDA, 2017).

Geographic trends in protein concentration have been observed. Soybean grown in northern and western soybean-producing area tend to have 1.5-2% less protein than those grown in southern states where the growing season is warmer and longer (Hurburgh, 1990). Yaklich et al. (2002) compared the protein and oil concentrations from the Northern and Southern Regional Uniform Tests from 1950-1998. Overall mean protein and oil concentrations were higher in the Southern Region (182 g kg^{-1} oil and 358 g kg^{-1} protein on a 13% moisture basis) compared to the Northern Region (179 g kg^{-1} oil and 354 g kg^{-1} protein).

Breeding for increased protein concentration

Protein concentration is a highly heritable trait with h^2 between 0.56 and 0.92 reported (Brummer et al., 1997). High heritabilities coupled with genetic diversity make high protein concentration a good candidate for breeding. The trait can also be phenotyped relatively quickly using a near-infrared reflectance (NIR) analyzer. Several breeding methods have been employed and have been successful at producing high protein lines; however, the negative correlation of this trait with yield has been difficult to overcome.

Recurrent selection has been used to increase protein concentration in soybean populations. In a study by Brim and Burton (1979), two populations, one derived from a cross between two highly adapted experimental lines with different maturities, protein concentrations, and oil concentrations and one derived from backcrossing nine unadapted high protein lines to an elite parent, were both subdivided in two for a total of four populations and underwent six cycles of recurrent selection. Protein concentration was increased and oil concentration decreased after four cycles of selection in all populations. Yield in the two derived from the backcrosses was significantly lower at C_6 than in C_0; however, in the other two populations, one population did not have a significant change and one had a yield increase. The population with a protein and yield increase demonstrated the ability to create high protein, high yield lines. Wilcox et al. (1998) also observed an increase in protein and decrease in oil after eight cycles of recurrent selection. Recurrent selection has been effective at increasing protein concentration, but it is

unrealistic method to be used on a large scale in a breeding program due to the difficulty in crossing soybeans.

Backcrossing is a technique that has shown promise to increase soybean protein concentration while maintaining yield. Wehrmann backcrossed "Pando", a high protein and low yielding plant introduction, to three high-yielding recurrent parents representing the maturity groups grown in Iowa to form BC_2F_2 lines. None of the lines had protein concentration higher than Pando, but 72% had a higher concentration than the recurrent parent and 19% were not significantly different than the donor parent. Approximately 19% of the BC_2F_2 lines yielded as much as the recurrent parent but with a higher protein concentration, demonstrating the ability to obtain high protein lines with good yields via backcrossing (Wehrmann et al., 1987). Wilcox and Cavins (1995) also used Pando as a donor parent in a study evaluating protein and yield. They selected for high protein and yield between each backcross and were able to identify a BC_3 line that combined high protein concentration with yield greater than one of the controls.

In the previously mentioned Recker et al study, 26 generations of random mating resulted in non-significant genetic correlations between yield and protein (2014). This suggests there can be success in increasing yield and protein simultaneously when the appropriate breeding strategy is implemented.

Marker Assisted Breeding and QTLs

Marker assisted breeding shows promise in aiding the development of high protein, high yielding soybean cultivars. Over 147 marker associations with protein concentration have been mapped to all soybean chromosomes. They are listed at: http://www.soybase.org/search/index.php?qtl=protein (Soybase, 2017). Three QTL associated with increased protein have been confirmed according to the rules of the soybean genetics community (Soybase, 2017).

In 1992, Diers et al. (1992) used 243 restriction fragment length polymorphism (RFLP) markers and identified two QTL controlling protein and oil concentration in a population of $F_{2:3}$ lines derived from a cross between the *G. max* experimental line, A81-356022, and the *G. soja* plant introduction, PI468916. The QTL were mapped to chromosome (chr) 20 (formally linkage group (LG) I) and chr 15 (LG E) with the *G. soja* alleles associated with increased protein

concentration. Following the Diers study, Sebolt et al. (2000) retested the QTL identified on chr 20 and chr 15. They backcrossed the two *G. soja* QTL alleles into the A81-356022 background to study the allelic effects on protein and oil concentration and agronomic traits. In the second part of the study, they developed three BC_3 populations to test the *G. soja* chr 20 QTL allele in different genetic backgrounds. Markers linked to the QTL on chr 15 were not associated with seed composition or agronomic traits, but markers linked to QTL allele from *G. soja* on chr 20 were associated with higher protein concentration in the first part of the study and in two out of the three BC_3 populations. Additionally, lines with the *G. soja* chr 20 QTL allele had reduced yield, smaller seeds, increased height, and later maturity.

Lee et al. (1996) evaluated two biparental populations, "Young" x PI416937 and PI97100 x "Coker 237," with RFLP markers to identify QTL associated with protein and oil concentration. A QTL on chr 15 was associated with increased protein in both populations; however, the other four QTL identified were population specific. Researchers sought to confirm the QTL from the Lee study by using the same RFLP markers in an independent PI97100 x Coker 237 population. Also, the original RFLP markers plus additional simple sequence repeat (SSR) markers in the same genetic regions were used in an independent Young x PI416937 population. The populations were grown in multiple environments. In the PI97100 x Coker 237 population, two out of four previously described protein QTL were confirmed including the one on chr 15; however, none of the previously reported protein QTL were confirmed in the Young x PI416937 population. This study proved that while it can be difficult to validate QTL in several genetic backgrounds, it is a necessity before selecting for QTL in breeding programs (Fasoula et al., 2004).

With rapid advances in genotyping technology and statistical modeling, association mapping has become a powerful tool to map protein QTL. Additionally, high throughput genotyping platforms have exponentially increased the number of markers used in mapping studies. While improved technology has allowed for more accurate and precise QTL mapping, the chr 15 and chr 20 QTL loci originally identified by Diers et al (1992) are still detected at a high frequency across mapping populations and appear to have the largest effect on protein concentration (Bolon et al., 2010; Hwang et al., 2014, Vaughn et al., 2014; Bandillo et al., 2015; Phansak et al., 2016).

Overview of QTL inconsistencies in mapping studies

The Fasoula et al. (2004) study supports the commonly observed trend of inconsistent QTL detection. Genetic background, heritability, population size, number of environments, and number and density of genetic markers all play a major role in QTL detection (Kao and Zeng, 1997). The main reasons for QTL inconsistency are experimental design and low detection power. Beavis (1994) designed a simulation study to examine QTL detection and the inconsistency of QTL analysis. Phenotypic variances associated with correctly identified QTL were greatly overestimated when progeny number was small (n=100) and but were more accurate when progeny number was large (1000). The highest power to detect QTL occurs when there is a small number of true QTL, a large number of progeny, and high heritabilities. Limitations to these parts of a QTL study, namely population size, prevent the detection of true QTL. Other possibilities for inconsistent QTL detection and confirmation can be attributed to allelic relationships and epistasis between a QTL and genetic background (Blanc et al., 2006).

Fine mapping and candidate genes for protein concentration

In order to develop more efficient markers for marker assisted selection, it is important to fine map a validated QTL. Fine mapping allows for the identification of better markers that are closer to the region of interest making tasks such as screening germplasm and gene cloning easier. The QTL on chr 20 originally identified in the aforementioned Diers study is currently one of the best characterized QTL associated with protein concentration. It has been mapped to the same region in several populations using different high protein parents which suggests there could be different alleles at this locus or the same allele in many sources (Diers et al., 1992; Brummer et al., 1997; Chung et al., 2003; Sebolt et al., 2003; Warrington et al., 2015). Nichols et al. (2006) localized the QTL to a 3-cM interval between SSR marker Satt239 and amplified fragment length polymorphism (AFLP), ACG9b. An evaluation of agronomic traits in lines with this QTL was inconsistent. This fine mapping study confirmed the QTL in the two backcross populations.

With recent advances in genomics and the availability of the soybean genome sequence, the QTL on chr 20 has further been characterized. Bolon et al. (2010) utilized transcript

profiling via Affymetric Soy GeneChip microarray and Illumina high-throughput transcriptome sequencing of a near isogenic line (NIL) pair differing for the high and low protein alleles for the chr 20 QTL. Fourteen genes mapped to the same region on chr 20 as the protein QTL in the Nichols et al. (2006) study, and 13 out of the 14 genes had transcript levels that were differentially expressed between the two NIL genotypes. These represent possible candidate genes involved in the determination of protein seed concentration.

Many QTL have been identified; however, very few have been validated, evaluated for agronomic traits, or fine-mapped. If QTL are going to be useful in a breeding program, they must be better characterized.

Sudden Death Syndrome

Sudden death syndrome (SDS) was first observed in Arkansas in 1972 by H.J. Walters (Hirrel, 1983). Since that time, the disease has spread to almost all soybean growing regions in the US and as far north as Canada (Rupe et al., 2001). In addition to multiple other countries, the disease also has been observed in Argentina and Brazil, the major soybean growing countries in South America (Nakajima et al., 1993; Ploper, 1993; Chehri et al., 2014; Tewoldemedhin et al., 2015). Symptoms were first reported in Illinois in 1986 (Roy et al., 1997). From 1996 to 2009, SDS was listed as one of the most important soybean diseases in the US (Wrather et al., 2009; Wrather et al., 2010). Yield losses up to 80 percent have been attributed to the disease with yield losses between 5 percent and 15 percent more common (Roy et al., 1997). In 2009, estimated yield loss due to SDS was over 34.4 million bushels/0.94 million metric tons in the US (Wrather et al., 2010).

Fusarium virguliforme

The fungal causal agent of the disease was originally referred to as *Fusarium solani* (Mart.) Sacc. (Roy et al., 1989; Rupe, 1989) but in subsequent years was referred to as *F. solani* f. sp. *glycines* to designate its soybean host (Roy et al., 1997). In 2003, Aoki et al. (2003) observed two morphologically and phylogenetically distinct species classified as *Fusarium solani* which caused sudden death syndrome, *F. tucumaniae* in South America and *F.*

virguliforme in North America. *F. tucumaniae* is different from other species of *F. solani* in that it has longer and more slender footed conidia. *F. virguliforme* produces comma-shaped sporodochial conidia in addition to septate, falcate, footed aerial conidia which differentiates it from *Fusarium phaseoli* (Aoki et al., 2003). In contrast to *F. tucumaniae*, *F. virguliforme* does not form perithecia (fruiting body). *F. virguliforme* has no known teleomorph (sexual reproductive stage) (Covert et al., 2007). Most of the literature now refers to the causal agent of sudden death syndrome in the US as *F. virguliforme* (Aoki and O'Donnell) (Aoki et al., 2003). In 2005, Aoki et al. identified two other species that caused SDS symptoms, *Fusarium brasilense* and *Fusarium cuneirostrum* which differ morphologically by their sporodocial conidia and phylogenetically (Aoki et al., 2005). It was later determined that *Fusarium cuneirostrum* did not cause SDS (O'Donnell et al., 2010; Aoki et al., 2012). An additional causal agent of SDS from Argentina and Brazil has been identified more recently, *Fusarium crassistipitatum* (Aoki et al., 2012).

Fusarium virguliforme overwinters freely in the soil and in plant residue as chlamydospores. It infects roots early in the growing season when the fungus penetrates the root-cap zone and then spreads intercellularly through the corticular tissue including the xylem and phloem by its hyphae (Navi and Yang, 2006). Rupe et al. (1999) found the soil concentration of *F. virguliforme* to be the highest in the top 15 cm of soil which is also were the most soybean roots are found.

In 1989, Roy et al. (1989) and Rupe (1989) completed Koch's postulates for SDS in two separate experiments growing and identifying the pathogen plated on potato dextrose agar (PDA). The isolates from plants displaying symptoms of SDS produced slow growing blue masses of macroconidia on PDA. There were few microconidia produced, and the PDA stained a dark maroon (Roy et al., 1989; Rupe, 1989). *F. virguliforme* was only isolated from the roots and lower stems and not from the leaves (Rupe, 1989). It was later observed that the fungus produces phytotoxins which are translocated to the leaves (Baker and Nemec, 1994; Jin et al., 1996; Ji et al., 2006). Four phytotoxins have been identified in cultures of *F. virguliforme*. One is a low weight, monorden identified by Baker and Nemec (1994), and another is a 17 kDa proteinaceous phytoxin identified by Jin et al. (1996). Additional work involving these two phytotoxins suggested the importance of light in the development of SDS symptoms. Light initiates a series of events where the free radicals develop, and the Rubisco large subunit is degraded initiating

programmed cell death (Ji et al., 2006). The identified third phytoxin is a 13.5-kDa acidic protein, FvTox1 reported by Brar et al. (2011). It was proposed FvTox1 is a single gene which produces free radicals that interrupt photosynthesis and is a major virulence factor in foliar symptom development of SDS (Brar et al., 2011; Pudake et al., 2013). Recently, a phytotoxic effector called FvNIS1 (necrosis-inducing secreted protein 1) has also been identified to be associated with SDS foliar symptoms (Chang et al., 2016).

Over 262 *F. virguliforme* isolates have been identified and genotyped although no races have been reported (O'Donnell et al., 2010). These isolates have been collected from soybean growing areas in the US, Canada, Argentina, and Brazil. Isolate FSG1 (Mont-1) is commonly used by the soybean research community. *F. virgilforme* isolates have different levels of aggressiveness and produce varying foliar and root symptoms. It is possible that some isolates are better at colonizing root tissue whereas others are better at translocating phytotoxin to the leaves (Li et al., 2009).

SDS symptoms

Foliar symptoms of SDS commonly develop when the plant is in its reproductive phase, namely after flowering and before pod fill (Hartman et al., 2015a). Early symptoms tend to appear on the uppermost leaves as small scattered, interveinal light green or chlorotic spots giving a mottled appearance. The spots then enlarge and can become necrotic or may run together to form larger areas of interveinal leaf cholorosis. If symptoms progress, most of the affected tissue will become necrotic with green tissue only remaining near the veins. In severe cases, defoliation occurs with petioles remaining attached to the stem. Flowers and pods can also be aborted with the younger ones aborted first. This leads to decreased seed and pod fill in addition to fewer pods (Roy et al., 1997). Field disease severity is usually rated on a 1 to 9 scale as described by Njiti et al. (1996).

Root symptoms become more pronounced with increased severity of foliar symptoms. Roots of infected plants can exhibit crown necrosis and lateral root rot. Inside the root, grayish to reddish brown discoloration radiates out from the pith; however, the pith remains white. Bluish sporulation may be seen on the taproot and lower stem. Plants are easily pulled from the soil (Roy et al., 1997).

Environmental conditions favoring SDS

From its earliest detection, SDS has been commonly associated with high soil moisture and cool temperatures. Irrigated fields with high fertility and high yield potential exhibit the worst symptoms. Scherm and Yang (1996) conducted greenhouse and field experiments examining the effect of temperature and moisture on the development of SDS symptoms. In the greenhouse, root symptoms were most severe at 18°C whereas foliar symptoms were most severe between 22°C and 24°C. Symptoms were light over 30°C. Favorable conditions for disease development in the roots followed by favorable conditions for plant development leads to increased disease. Disease increases because increased levels of the toxin in the roots are able to be translocated to the leaves better which leads to foliar symptoms. The wettest treatments in both greenhouses and fields produced the worst SDS symptoms. Compacted field areas that retain more moisture for longer period of time also show more symptoms than non-compacted areas (Scherm and Yang, 1996). Field irrigation during late to mid reproductive stages showed greater increases in SDS development compared to irrigation during vegetative stages (de Farias Neto et al., 2006). Soils amended with calcium phosphate, potassium phosphate, potassium sulfate, sodium phosphate, or potassium nitrate also resulted in a 21% to 45% increase in SDS severity (Sanogo and Yang, 2001).

Managing SDS

It is difficult to manage SDS. Foliar fungicides have limited effects on the control of *F. virguiliforme* as inoculation takes place underground in the root early in the growing season (Henricksen and Elen, 2005). A seed treatment labeled for SDS control was recently made commercially available (Hartman et al., 2015b). Results from industry tests appear promising; however, the treatment is recommended for use with resistant varieties.

Although there are rotations that decrease the presence of other soybean pathogens, there does not appear to be a rotation that significantly lowers the level of *F. virguliforme* in the soil (Rupe et al., 1997). Crop rotation is not only limited by the ability to of the pathogen to overwinter via chlamydospores but also its wide host range (Hartman et al., 2015b). *F.*

virguliforme symptomatic hosts include alfalfa (*Medicago sativa*), red clover (*Trifolium pratense*), common bean (*Phaseolus vulgaris*), pea (*Pisum sativa*), Canadian milk vetch (*Astragalus canadensis*), sugar beet (*Beta vulgaris*), and canola (*Brassica napus*) whereas asymptomatic hosts include corn (*Zea mays*), wheat (*Triticum aestivum*), ryegrass (*Lolium perenne* L.), pigweed (*Amaranthus palmeri*), and lambsquarters (*Chenopodium album*) (Kolander et al., 2012).

Tillage has been suggested as a cultural control for SDS (Hartman et al., 2015b). Wrather et al. (1995) found no-till increased SDS foliar symptoms when compared to disk-till and ridge-till. Vick et al. (2003) reported subsoiling greatly reduced foliar symptoms compared to no-till in compacted areas (Vick et al., 2003). The efficacy of tillage to control SDS is likely dependent on location and is likely best suited for control in fields with compaction problems.

Later planting dates have also been associated with decreased SDS symptoms. In addition to studying tillage, Wrather et al. (2005) observed the effect of mid-May, mid-June, and late July planting dates. In findings similar to Hershman et al. (1990), early planting resulted in increased SDS symptoms. However, later planting dates tend to lead to decreased yields, so this is not necessarily an appropriate control for farmers to use.

Resistant cultivars are the one of the best methods of SDS control. Field resistance has been classified as horizontal, rate reducing and partial. It is controlled by many genes and with high heritabilities reported (see below) (Njiti et al., 1996). Several resistant genotypes have been identified and have been made available to producers; however, none of the current resistance is complete. All genotypes will display some SDS symptoms if conditions are favorable (Hershman, 1990; Iqbal et al, 2001; Mueller et al., 2003).

SDS and SCN

In early observations of sudden death syndrome, Hirrel (1983) noted the presence of soybean cyst nematodes in affected fields. He noted SCN was associated with 70 to 80% of plants displaying SDS symptoms in 30 fields across four states (Hirrel, 1983).

McLean and Lawrence (1995) studied the effect of *F. virguliforme* on SCN in a coinoculation study. After 40 days, the fungus was found in 37% of examined cysts. It was also isolated from the cortex and syncytia in plant tissue near developing juveniles. Wounding caused

by the nematode can provide an opportunity for *F. virguliforme* to enter plant tissue. Roy et al. (2000) isolated *F. virguliforme* in cysts from a majority of fields in the Midwest and South affected with both pathogens (Roy et al., 2000). The ability of *F. virguliforme* to survive in SCN cysts can possibly influence the severity of SDS and its dispersal. The cysts provide a more protected environment from other soil microorganisms than either soil or soybean residue. In an isolated environment such as the cyst, there is little competition with other organisms leading to enhanced survival. Cysts are possibly also a food source to spur the growth of chlamydospores in the spring. The spread of SDS from its original place of detection in Arkansas could be in part due to *F. virguliforme* colonized cysts (McLean and Lawrence, 1995).

In early greenhouse studies where soil was inoculated with *F. virguliforme* and SCN, SCN led to more severe foliar SDS symptoms but was not necessary for disease infection (Roy et al., 1989). A field microplot study produced similar results. Foliar disease symptoms appeared three to seven days earlier and were more severe in plots with *F. virguliforme* and SCN than those with only *F. virguliforme* (McLean and Lawrence, 1993). In a later microplot study, *F. virguliforme* and SCN also damaged plants synergistically in coinoculated plots. Plots with natural levels of SCN also had high levels of SDS. In one season, only coinoculated plots displayed SDS symptoms. These results pointed to a positive correlation between the pathogens (Xing and Westphal, 2006). Enhanced symptoms of SDS caused by SCN may explain the typical scattered clustered pattern of diseased plants in a field. SCN is usually not distributed evenly through the field, so "hot spots" of SDS may occur where SCN populations are particularly high (Scherm and Yang, 1996).

Some studies have supported the association between SCN and SDS, but others have not. A correlation between cyst counts and disease severity in a Scherm et al. (1998) study on soil variables in SDS fields in Iowa was visible but weak, and the cross-correlation coefficients for SCN cysts were not always significant (Scherm et al., 1998). Gao et al. (2006) inoculated a susceptible genotype with different levels of *F. virguliforme* and SCN and conducted real-time polymerase chain reaction. Infection of soybean roots by SCN did not affect colonization by the fungus, and the only significant main effect was fungal population. Both pathogens reduced plant growth, but SCN did not increase foliar symptoms of SDS. Overall, statistical interactions between SCN and *F. virguliforme* were rarely significant.

Breeding for SDS resistance

Heritabilities for foliar disease symptoms of SDS of approximately 0.83 have been reported (Hnetkovsky et al., 1996; Njiti et al, 1996; de Farias Neto et al., 2007; Wen et al., 2014). A breeder can select for a highly heritable trait, so there is opportunity to improve SDS resistance through breeding.

Cultivars resistant to SDS have been identified in field and greenhouse studies (Lightfoot, 2015). Field screens have been conducted at locations with a history of SDS symptoms or locations inoculated with the pathogen (Chang et al., 1996; Hnetkovsky et al., 1996; de Farias Neto et al., 2006). Disease severity and disease incidence (DI) are taken about the R6 growth stage to calculate a disease index (DI) (Njiti et al., 1996). In greenhouse screenings, seeds are planted into soil with incorporated or layered inoculum. Seedlings are rated for foliar symptoms approximately 21 days after planting (de Farias Neto et al, 2008; Hartman et al., 1997; Hashmi et al., 2005). The field and greenhouse screening methods for SDS have pros and cons; however, they both have displayed success in identifying resistance.

There are several sources of resistance to SDS in soybean germplasm. Between 1997 and 2003, over 6800 plant introductions and cultivars in the USDA Soybean Germplasm Collection were screened for SDS resistance in a greenhouse (Hartman et al., 1997; Mueller et al., 2002; Mueller et al, 2003). Ninety lines were identified as having a moderate resistance level better than or equal to that of the check(s) in the study. These studies were beneficial in that they pinpointed germplasm that can be used to identify new resistance genes and alleles.

Marker-assisted breeding and SDS QTLs

Marker-assisted selection can make the breeding process more efficient by eliminating some laborious phenotyping by making selection for resistance more accurate. Currently, 56 marker associations with SDS resistance on 11 chromosomes have been identified and listed at: http://www.soybase.org/search/index.php?qtl=SDS (Soybase, 2017). Further research has indicated that QTL associated with resistance control either foliar symptom severity, root symptom severity, or both (Njiti et al., 1998; Triwotayakorn et al., 2005; Kazi et al., 2007; Abdelmajid et

al, 2012). Locations of several QTL have been the same across mapping populations (Njiti et al., 2002; Luckew et al., 2013; Wen et al., 2014).

Several QTL associated with SDS resistance have been identified, but only a few have been confirmed. An outline set forth by the Soybean Genetics Committee for confirming QTL can be found at: www.soybase.org. Njiti et al (1998) developed a secondary population of NILs from a recombinant inbred line (RIL) from the "Essex" x "Forrest" population that was used to map SDS QTL. The NILs were segregating within the regions of chr 18 (LG G) and chr 6 (LG C2) where QTL associated with SDS resistance had been previously mapped in the primary RIL population. The Forrest allele of the region on chr 18 was associated with a decrease in disease index (DI) and infection severity (DS) across locations with p=0.0004. de Farias Neto (2007) mapped resistance QTL on chr 19 (LG L) and chr 17 (LG D2). In a "Ripley" x "Spencer" population, the QTL on chr 17 mapped to the same genetic region as a QTL they also mapped in a population developed by crossing PI567374 x "Omaha." The QTL was then tested in an F_2 population derived from a BC_1F_2 with PI567374 as the donor parent and Omaha as the recurrent parent. It was also tested in an F_8 population developed from an F_5-derived line from the original Ripley x Spencer mapping population that was segregating for the region of interest. In both populations, there was a significant association between markers linked to the QTL and greenhouse DS. Plants with the Ripley allele or the PI567374 allele had less foliar symptoms. The QTL on chr 17 was designated cqSDS-001 (de Farias Neto et al., 2007).

While many QTL for SDS resistance have been come mapped in bi-parental populations, recently association mapping has become a useful tool to map SDS resistance QTL (Lightfoot, 2015). These studies have used diverse mapping panels to mapped novel loci for resistance along with previously mapped SDS QTL. (Wen et al. 2014; Chang et al., 2016).

Fine mapping and candidate genes for SDS resistance

One of the best characterized genetic regions associated with SDS resistance is a cluster of loci on chr 18. Meksem et al. (1999) examined this genomic region in a fine mapping study with resistance derived from Forrest. SCN parasitism, SDS foliar symptoms, and SDS root symptoms were rated in NILs with recombinations within the region of interest. The region near the RFLP marker Bng122D was significantly associated with SDS foliar and root symptoms

whereas Satt309 was significantly associated with resistance to SCN and SDS foliar symptoms but not with root symptoms. The latter foliar resistance gene could not be separated from *rhg1*. Although putative locations of regions related to SDS resistance were reported, due to low map resolution they were unable to determine whether resistance to SDS and SCN in Forrest was conferred by a cluster of four genes with two closely linked pairs or two genes with pleiotrophic effects where one controlled SDS root and foliar symptoms and one controlled SCN infection and SDS foliar symptoms. Triwitayakorn (2005) also examined the cluster of genes on chr 18 associated with SDS resistance, and identified candidate genes for *QRfs1* and *QRfs2*. *QRfs2* clustered with *rhg1* and provided resistance to foliar symptoms in the Meksem et al. study whereas *QRfs1* provided resistance to root symptoms. A 0.2 cM resolution map of the region on chr 18 was developed between the 7.5 cM interval between Satt309 and SIUC-Sat122 (Triwitayakorn et al., 2005). A gene within the *Rfs2/rhg1*-a locus, *GmRLK18-1*, was later inserted as a transgene and resistance to SDS and SCN was observed (Srour et al., 2012).

More genetic studies to confirm and fine map QTL are needed for increasing the efficiency of incorporating SDS resistance into cultivars by breeders because these studies will increase the predictability of markers linked to resistance traits.

Soybean Cyst Nematode

Soybean cyst nematode (SCN) or *Heterodera glycines* Ichinohe is the most economically important pathogen of soybean in the US. From 2006-2009, the estimated average annual loss to SCN was 128.6 million bushels/3.5 million metric tons representing a $1.286 billion loss to producers (Koenning and Wrather, 2010). SCN was first identified in North Carolina in 1953; however, since this time, the pathogen has spread to almost all major soybean producing areas.

SCN is a soilborne pathogen with an egg and four juvenile stages. The eggs can overwinter within a female cyst under harsh environmental conditions (Alston and Schmitt, 1988; Jackson et al., 2005). The first stage juvenile develops within an egg and molts to become a second stage/infective stage juvenile (J2). The J2 moves a short distance through the soil to the root tips. It penetrates the root and establishes a feeding site called a syncytium and engorges. The juvenile molts three more times before becoming an adult. Females become immobile and continue to feed on the root. Their bodies swell and become yellow, lemon-shaped cysts that

protrude from the root and contain approximately 100 to 200 eggs. Eggs only develop if they are fertilized by a male. At death, the cysts are brown and dislodge from the root. Adult males are vermiform and mobile. They stop feeding, exit the root to fertilize females, and die. The life cycle takes about 25 to 40 days with several generations occurring in a single growing season (Triantaphyllou and Hirshmann, 1962; Jardine and Todd, 2001).

SCN symptoms

The most common symptom of SCN is yield loss; so the producer usually does not become aware of an SCN problem until it is too late for yield to be protected. Visible symptoms include stunting, chlorosis, decreased root mass, and/or decreased nodulation. In severe cases, premature plant death occurs. Many of these symptoms are indistinguishable from other diseases and abiotic stresses. Soil sampling to identify cysts is the best way to confirm if a symptom is caused by SCN (Jardine and Todd, 2001).

Conditions conducive to SCN development

SCN is heavily influenced by environmental conditions including soil type and moisture levels. Greater soil and root populations are observed in sandy soils compared to clay soils (Todd and Pearson, 1988). Levels of SCN can increase to damaging levels in fine textured soils but at a lower rate due to decreased reproductive levels. Several stages of the nematode's life cycle including hatching, movement, and development require aerobic respiration. Fine textured soils retain water for longer periods of time creating anaerobic conditions unfavorable to the nematode's survival. Heavier textured soils can allow for easier nematode movement as there is more space between soil particles (Koenning and Barker, 1995).

SCN survival is also affected by temperature. Slack et al. (1972) noted larvae survived for over 630 days in water at temperatures between 0°C and 12°C but died when ice crystals formed within close proximity or after a day at 40°C. In natural soils, nematodes survived 6 to 8 years between 0°C and 20°C and were not immediately killed by extreme high (above 40°C) and low temperatures (below freezing). At temperatures over 20°C larvae survival time decreased with increased temperatures. Optimum temperature for an egg to hatch is 24°C where

there is low mortality and development is rapid. Hatching has been observed between 20°C and 30°C (Alston and Schmitt, 1988). Penetration, development, and reproduction of SCN are negatively affected at temperatures below 14°C and above 33°C (Hamblen et al., 1972).

Control methods for SCN

Rotation to non-host crops such as corn, wheat, or sorghum can be an effective means of control, but these methods do not completely eradicate the pathogen from the soil because of an egg's ability to overwinter in the soil protected within the cyst for several years (Jackson et al., 2005; Miller et al., 2006; Porter et al., 2001). Multiple seed treatments have also been made available to producers, but they are recommended to be used in conjunction with genetic resistance. The best method of SCN control is resistant cultivars. SCN are able to penetrate the root and form syncytia in SCN-resistant cultivars; however, the syncytia either form slowly or become necrotic soon after they are formed. This causes the nematode to starve to death (Williamson and Hussey, 1996).

Genetic resistance to SCN

Soybean genetic resistance to SCN was identified shortly after the pathogen was found in the US. Caldwell et al. (1960) used a traditional genetic study and showed that resistance from PI548402 ("Peking") was controlled by the three recessive and independent alleles: *rhg1*, *rhg2*, and *rhg3*. In a later study, Matson and Williams (1965) reported a fourth dominant resistance gene from Peking that was named *Rhg4*. Rao-Arelli et al. (1992) reported another dominant allele in PI88788 which was given the designation *Rhg5* (Rao-Arelli et al., 1992).

Over 118 soybean plant introductions have been identified as SCN resistant and represent potential sources of resistance (Arelli et al., 2000). It appears genetic diversity for SCN resistance exists; however, it is narrow with only seven sources of resistance currently used by breeders to create commercial cultivars. Several of the commonly used resistance sources have one or more genes in common making genetic diversity even more limited. Most SCN resistance in cultivars is derived from PI88788 (Diers and Arelli,, 1999; Shier, 2008). This PI has shown a broad level of resistance to HG types, and it is currently the source of resistance that has been

most successfully used in developing cultivars with the best agronomic performance. Of 760 maturity group I to IV cultivars listed as having SCN resistance, 705 had resistance derived from PI88788 alone (Concibido et al., 2004).

With the heavy reliance of breeders on just a few sources of resistance, it is important to continue to search for novel resistance sources. This need is even more important when the ability of nematodes to adapt and overcome plant resistance is taken into account. In a survey of soil samples from Illinois, 70% of the SCN populations found could overcome PI88788 resistance (Niblack et al., 2008). One way to find new resistance genes is by screening soybean relatives such as *Glycine soja*. It has been shown that stacking resistance alleles from wild soybean with those from domestic sources can increase resistance (Kim et al., 2011).

rhg1

The *rhg1* locus has been mapped in almost all SCN the resistant accessions used in mapping studies and is the most common SCN resistance gene in cultivars in the US. Brucker et al. (2005) identified allelic differences in a population segregating for *rhg1* alleles from PI88788 and Peking. The resistance allele from PI88788 is now designated *rhg1-b*.

SCN resistant cultivars protect yields under high SCN pressure, but they were shown to yield 5 to 10% less than susceptible cultivars when pressure is low (Noel, 1992). Mudge et al. (1996) tested two independent breeding populations with *rhg1* from PI 209332. *Rhg1* was shown to be linked to two QTL conferring yield depression. They proposed this may be due to yield drag where yield is suppressed by pleiotrophic effects of resistance genes or due to the effect of yield reducing alleles in coupling linkage with the resistance gene. Kopisch-Obuch et al. (2005) tested five soybean populations of NILs segregating for *rhg1-b*. They detected yield depression in one of the NIL populations but suggested further research to confirm and refine the findings.

Classifying SCN isolates

The SCN HG type system is the current way to classify SCN populations. Prior to this system, resistance was classified using the SCN race system. This method incorporated four sources of resistance: “Pickett”, Peking, PI88788, and PI90763. If a SCN race had a female

index (FI) above 10 percent it was designated a "+"whereas a female index at or below 10 percent was designated a "-" (Schmitt and Shannon, 1992). Female index is calculated by dividing the mean number females on a test soybean line by the mean number of females on the standard susceptible and multiplying by 100. The problem with the race classification system was that it only incorporated four sources of resistance and did not adequately characterize SCN's diversity. Additionally, the term "race" was not appropriate because the genotype of a nematode population could not be inferred from the test. The HG type system addresses some of these problems by allowing for the addition of several sources of resistance. SCN populations are collected by environment and are tested on several differentials (Peking, PI88788, PI90763, PI427654, PI209332, PI89772, and "Cloud") with each FI noted. A standard susceptible, "Lee 74", is used to calculate FI. A FI above ten receives a numerical designation, and the HG type corresponds to the differentials on which a nematode population can reproduce (Niblack et al., 2002).

Breeding for SCN resistance

Breeding for SCN resistance is difficult. The pathogen is variable, and most resistance sources control only a subset of HG types. Often, several genes are needed to produce broad based resistance. The sources of resistance also are not adapted to the soybean growing areas in US and can be hard to introduce into elite germplasm (Concibido et al., 2004). The recent sequencing of the soybean genome combined with the exponential decrease in sequencing costs has opened the doorway for marker assisted selection (MAS) to become a more viable option in breeding for SCN resistance (Concibido et al., 2004; Schmutz et al., 2010). The conventional way to screen for SCN resistance is to inoculate plants in a greenhouse and calculate their female indices; however, molecular technology presents another option. Once markers linked to QTL associated with resistance are identified, they can then be used to select resistant and susceptible genotypes. Marker-assisted selection for SCN resistance is of great interest to breeders because the conventional greenhouse screening is expensive, time consuming, and sometimes unreliable. Markers have the ability to address each of these flaws to produce cheaper, more accurate data in a shorter amount of time. Additionally, markers also present the opportunity to identify new resistance genes because genotypes can be easily screened with thousands of markers.

QTLs for SCN resistance

From 1986 to 2006, over 17 papers have reported over 62 marker-based quantitative trait locus (QTL) associations with SCN resistance (Guo et al., 2006). In 1994, Concibido et al. detected a major resistance QTL on chr18 using a population developed from a cross between SCN-susceptible experimental line M83-15 and the line M85-1430, which has SCN resistance derived from PI209332. This QTL was tightly linked to the RFLP marker K069 and had an R^2 value of 0.16 (Concibido et al., 1994). The QTL was later assigned the gene symbol *rhg1*. Since this study, QTL in the *rhg1* interval have been identified in multiple backgrounds (Guo et al., 2006; Wu et al, 2009; Vuong et al., 2011).

Seven SCN resistance QTL have been confirmed based on rules from the Soybean Genetics Committee (Soybase, 2017) including *rhg1*, which has the confirmed QTL designation cqSCN-001 and *Rhg4*, which has the confirmed QTL designation cqSCN-002. Kim et al. (2010) was able to fine map the *rhg1-b* allele from PI88788 by testing lines with known genetic recombination within the area of the *rhg1* locus. Lines were phenotyped for resistance to the SCN isolate PA3 (HG type 0). Analysis placed *rhg1-b* within a 67-kb region between the SSR markers, BARCSOYSSR_18_0090 and BARCSOYSSR_18_0094. This region was 34 kb from a receptor like-kinase gene for *rhg1* from "Peking" described by Lightfoot and Meksem (2002) (Kim et al., 2010).

Young et al (1999) identified PI567516C as the only source of resistance to LY1 nematodes, a highly virulent synthetic population derived from a mass mating of HG Type 1.2.3 females with HG Type 1.2 males. Further research confirmed the accessions's resistance (Arelli and Young, 2005) and its genetic distinction from another commonly used resistance source, "Hartwig" (Chen et al., 2006). QTL were mapped using SSR markers in a population derived from a PI567516C x Hartwig cross. A genetic region on chr 10 (LG O), defined by Satt592, Satt331, and Sat_274, was associated with LY1 resistance (Arelli et al., 2010). These markers can be used to aid in marker assisted selection for nematode resistance. Vuong et al. identified four QTL associated with SCN resistance using $F_{2:3}$ progeny from a Magellan x PI567516C cross. One of these QTL was in the marker interval Satt_038-Satt592 on chr 10. It was confirmed

in a $F_{6:7}$ RIL population derived from the same cross. The chr 10 QTL conferred resistance to HG types 2.5.7, 0, 2.7, 1.3.5.6.7, and LY1 (Vuong et al. 2010).

QTL have also been mapped in *G. soja* accessions. Wang et al. (2000) identified two significant QTL in populations derived from a cross between the *G. soja* PI468916 and *G. max* experimental line A81-356022. The QTL were mapped to chr 18 and chr 15 with the *G. soja* allele conferring resistance to SCN. Kabelka et al. (2005) were able to confirm both QTL in a population of $BC_4F_{3:4}$ lines developed with PI468916 as the donor parent and A81-356022 as the recurrent parent. A significant association was observed between the AFLP and SSR markers on chr 15 and chr 18 and the SCN phenotypes of the population ($P<0.05$). Additionally, AFLP markers and bulk segregant analysis identified additional markers in the regions containing the QTLs. The QTL on chr 18 was mapped to a 18.5 cM region with an average marker distance of 3.1 cM whereas the QTL on chr 15 was mapped to a 13.2 cM region with an average marker distance of 0.5 cM. The QTL on chr 15 and chr 18 were designated as cqSCN-006 and cqSCN-007, respectively.

Recently, Kim and Diers (2013) fine mapped cqSCN-006 and cqSCN-007. Over 1200 F_5 and F_6 plants were tested with SSR markers flanking cqSCN-006 on chr 15 to identify recombinants. The progeny from 20 recombinant plants were then tested with more markers and screened for SCN resistance in the greenhouse. This effort resulted in the mapping of the QTL to a 803.4 kb region. A similar strategy was used to fine map cqSCN-007 to a 146.5 kb region on chr 18. In this effort, over 1600, F_5 plants were tested with SSR markers flanking the QTL to identify recombinants. The progeny from 18 recombinant plants were then tested with additional markers and screened for SCN resistance in the greenhouse.

The identification and characterization of QTL associated with SCN resistance aids in more advanced genomic studies aimed at pinpointing candidate genes and their biological functions.

Candidate genes and resistance mechanisms

Advances in genomics have enabled researchers to build upon the information uncovered in previous genetic studies to learn more about SCN-resistance.

Liu et al. (2012) cloned *Rhg4* from Forrest (Peking resistance). They then conducted mutation analysis, gene silencing, and transgenic complementation to verify that the cloned gene conferred resistance. The gene encoded a serine hydroxymethytransferase (*SHMT*), an enzyme needed for the interconversion of serine and glycine and therefore cellular one-carbon metabolism. Resistant and susceptible alleles of *Rhg4* had two genetic polymorphisms that distinguished them. The polymorphisms disrupt the plants folate pathway which can trigger either a hypersensitive response or a resistance signaling pathway (Liu et al., 2012)

Cook et al. (2012) reported a novel mode of plant resistance conditioned by *rhg1-b*. The 31-kb segment at this locus encoded an amino acid transporter, an alpha SNAP protein, and a wound inducible domain protein. Moreover, susceptible genotypes had only one copy of this segment whereas resistance genotypes had ten copies. When the genes were overexpressed on an individual basis, there was no resistance; however, when the genes were overexpressed together, resistance was enhanced. This indicated copy number variation plays an important role in SCN resistance.

Genetic resistance summary

As one of few practical methods of SCN control, resistance is an important trait for producers. Breeders are actively working to release cultivars that protect yield from SCN damage. Genomic studies have revealed the complexity of SCN resistance and have also provided genetic information for breeders to exploit in creating new cultivars. Genetic gain has been increased with the ability to assemble beneficial genotypic variation. With genetic architecture of SCN resistance better characterized, heritability can also increase while selection intensity and cycle time can decrease.

Currently, QTL studies and new marker technology are having the greatest impact on resistance breeding. Lines containing SCN resistance genes such as *rhg1* can now be easily identified with markers. This allows selections to be made faster and more accurately. With advances in genetic mapping and marker technology occurring quickly, it may become easier to identify new and more effective sources of resistance.

References

Abdelmajid, K M., Ramos, L., Leandro, L., Mbofung, G., Hyten, D. L., Kantartzi, S. K., Grier, V., Robert, L., Njiti, V. N., and Meksem, K. 2012. The 'PI 438489B'by 'Hamilton'SNP-based genetic linkage map of soybean [*Glycine max* (L.) Merr.] identified quantitative trait loci that underlie seedling SDS resistance. J. Plant Genome Sci. 1:18-30.

Alston, D., and Schmitt, D. 1988. Development of *Heterodera Glycines* life stages as influenced by temperature. J. Nematol. 20:366-372.

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

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

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

Arelli, P. R., Sleper, D. A., Yue, P., and Wilcox, J. A. 2000. Soybean reaction to races 1 and 2 of *Heterodera glycines*. Crop Sci. 40:824-826.

Arelli, P. R., Concibido, V. C., and Young, L. D. 2010. QTLs associated with resistance in soybean PI567516C to synthetic nematode population infecting cv. Hartwig. J. Crop Sci. Biotech. 13:163-167.

Arelli, P. R., and Young, L. D. 2005. Genetics of resistance in soybean PI567516C to LY1 nematode population infecting cv. Hartwig. Crop Sci. Soc. Am. p 234 (Abstracts).

Bandillo, N., Jarquin, D., Song, Q., Nelson, R., Cregan, P., Specht, J., and Lorenz, A. 2015. A population structure and genome-wide association analysis on the USDA soybean germplasm collection. The Plant Genome. 8.

Baker, R., and Nemec, S. 1994. Soybean sudden death syndrome: Isolation and identification of a new phytotoxin from cultures of the causal agent, *Fusarium solani*. Phytopathology. 84: 1144.

Beavis, W. D., O. S. Smith, D. Grant, and R. Fincher. 1994. Identification of Quantitative Trait Loci Using a Small Sample of Topcrossed and F4 Progeny from Maize. Crop Sci. 34:882-896.

Bellaloui, N., and Mengistu, A. 2008. Seed composition is influenced by irrigation regimes and cultivar differences in soybean. Irrig. Sci. 26: 261-268.

Bernard, R. L., and Cremeens, C. R. 1988. Registration of 'Williams 82'soybean. Crop Sci. 28:1027-1028.

Blanc, G., Charcosset, A., Mangin, B., Gallais, A., and Moreau, L. 2006. Connected populations for detecting quantitative trait loci and testing for epistasis: An application in maize. Theor. Appl. Genet. 113: 206-224.

Bolon, Y., Joseph, B., Cannon, S. B., Graham, M. A., Diers, B. W., Farmer, A. D., May, G.D., Muehlbauer, G. J., Specht, J. E., Tu, Z. J., and Weeks, N. 2010. Complementary genetic and genomic approaches help characterize the linkage group I seed protein QTL in soybean. BMC Plant Biology. 10:41.

Borras, L., Slafer, G. A., and Otegui, M. 2004. Seed dry weight response to source sink manipulations in wheat, maize and soybean: A quantitative reappraisal. Field Crops Res. 86: 131-146.

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

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

Brucker, E., Carlson, S., Wright, E., Niblack, T., and Diers, B. 2005. *Rhg1* alleles from soybean PI 437654 and PI 88788 respond differentially to isolates of *Heterodera glycines* in the greenhouse. Theor. Appl. Genet. 111:44-49.

Brucker, E., Niblack, T., Kopisch-Obuch, F. J., and Diers, B. W. 2005. The effect of on reproduction of in the field and greenhouse and associated effects on agronomic traits. Crop Sci. 45:1721-1727.

Brummer, E. C., G. L. Graef, J. Orf, J. R. Wilcox, and R. C. Shoemaker. 1997. Mapping QTL for Seed Protein and Oil Content in Eight Soybean Populations. Crop Sci. 37:370-378.

Caldwell, B. E., Brim, C., and Ross, J. 1960. Inheritance of resistance of soybeans to the cyst nematode, *Heterodera glycines*. Agron. J. 52:635-636.

Carrera, C., Martínez, M. J., Dardanelli, J., and Balzarini, M. 2009. Water deficit effect on the relationship between temperature during the seed fill period and soybean seed oil and protein concentrations. Crop Sci. 49:990-998.

Carter, T. E., Burton, J. W., and Brim, C. A. 1986. Registration of NC 101 to NC 112 soybean germplasm lines contrasting in percent seed protein. Crop Sci. 26:841-842.

Carter T.E., Nelson R., Sneller, C. H., Cui Z. 2004. Genetic diversity in soybean. p. 303-416. In Soybeans: Improvement, Production, and Uses. Vol. 16. eds Boerma HR , Specht JE. ASA-CSA-CSSA, Madison, WI.

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

Chang, S., Doubler, T., Kilo, V., Suttner, R., Klein, J., Schmidt, M., Gibson, P., and Lightfoot, D. 1996. Two additional loci underlying durable field resistance to soybean sudden death syndrome (SDS). Crop Sci. 36:1684-1688.

Chehri, K., B. Salleh, and L. Zakaria. 2014. Fusarium virguliforme, a soybean sudden death syndrome fungus in Malaysian soil. *Australasian Plant Disease Notes, 9*:128.

Chen, Y., Wang, D., Arelli, P., Ebrahimi, M., and Nelson, R. L. 2006. Molecular marker diversity of SCN-resistant sources in soybean. Genome. 49:938-949

Chung, J., Babka, H. L., Graef, G. L., Staswick, P. E., Lee, D. J., Cregan, P. B., Shoemaker, R. C., and Specht, J. E. 2003. The seed protein, oil, and yield QTL on soybean linkage group I. Crop Sci. 43:1053-1067.

Clarke E., Wiseman J. 2000. Developments in plant breeding for improved nutritional quality of soya beans I. Protein and amino acid content. J. Agric. Sci (Cambridge). 134:111-124.

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

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

Cook, D. E., Lee, T. G., Guo, X., Melito, S., Wang, K., Bayless, A. M., Wang, J., Hughes, T. J., Clemente, T. E. and Diers, B. W. 2012. Copy number variation of multiple genes at *Rhg1* mediates nematode resistance in soybean. Science. 338:1206-1209.

Covert, S., Aoki, T., O'Donnell, K., Starkey, D., Holliday, A., Geiser, D., Cheung, F., Town, C., Strom, A., Juba, J., Scandiani, M, and Yang, X. B. 2007. Sexual reproduction in the soybean sudden death syndrome pathogen *Fusarium tucumaniae*. Fungal Genet. Biol. 44:799-807.

Cregan, P., Mudge, J., Fickus, E., Danesh, D., Denny, R., and Young, N. 1999. Two simple sequence repeat markers to select for soybean cyst nematode resistance conditioned by the *rhg1* locus. Theor. Appl. Genet. 99:811-818.

Cromwell, G.L. 2012. Soybean meal: An exceptional protein source. Soybean Meal InfoCenter; Ankeny, IA: available at <http://www.soymeal.org/factsheets/>. Accessed January 31, 2017.

Dardanelli, J. L., Balzarini, M., Martínez, M., Dardanelli, J. L., Balzarini, M., Martínez, M. J., Cuniberti, M., Resnik, S., Ramunda, S. F., Herrero, R., and Baigorri, H. 2006. Soybean maturity groups, environments, and their interaction define mega-environments for seed composition in Argentina. Crop Sci. 46:1939-1947.

de Farias Neto, Austeclinio L., Hartman, G. L., Pedersen, W. L., Li, S., Bollero, G. A., and Diers, B. W. 2006. Irrigation and inoculation treatments that increase the severity of soybean sudden death syndrome in the field. Crop Sci. 46:2547-2554.

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

Diers, B. W., and Arelli, P. R. 1999. Management of parasitic nematodes of soybean through genetic resistance. p. 4-7. In Proceedings of world soybean research conference, 6th (Kauffman HE), Chicago, IL.

Diers, B. W., Keim, P., Fehr, W. R., and Shoemaker, R. C. 1992. RFLP analysis of soybean seed protein and oil content. Theor. Appl. Genet. 83:608-612.

Dornbos, D., and Mullen, R. E. 1992. Soybean seed protein and oil contents and fatty acid composition adjustments by drought and temperature. J. Am. Oil Chem. Soc. 69:228-231. doi:10.1007/BF02635891

Farias Neto, A.L., Schmidt, M., Hartman, G. L., Shuxian, L., and Diers, B. W. 2008. Inoculation methods under greenhouse conditions for evaluating soybean resistance to sudden death syndrome. Pesquisa agropecuária Brasileira, 43:1475-1482.

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

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

Gibson, L. R., and Mullen, R. E. 1996. Soybean seed composition under high day and night growth temperatures. J. Am. Oil Chem. Soc. 73:733-737.

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

Guo, B., Sleper, D., Lu, P., Shannon, J., Nguyen, H., and Arelli, P. 2006. QTLs associated with resistance to soybean cyst nematode in soybean: Meta-analysis of QTL locations. Crop Sci. 46:595-602.

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

Hamblen, M., Slack, D., and Riggs, R. 1972. Temperature effects on penetration and reproduction of soybean-cyst nematode. Phytopathology. 62:762.

Hanson, W. D. 1991. Seed protein content and delivery of assimilates to soybean seed embryos. Crop Sci. 31:1600-1604.

Hanson, W. D., Leffel, R. C., and Howell, R. W. 1961. Genetic analysis of energy production in the soybean. Crop Sci. 1:121-126.

Hartman, G., Huang, Y., Nelson, R., and Noel, G. 1997. Germplasm evaluation of *Glycine max* for resistance to *Fusarium solani*, the causal organism of sudden death syndrome. Plant Dis. 81:515-518.

Hartman, G.L., Leandro, L.F., and Rupe, J.C. 2015. Sudden Death Syndrome." In: G.L. Hartman, J.C. Rupe, E.J. Sikora, L.L. Domier, J.A. Davis, and K.L. Steffy (eds), Compendium of soybean diseases (5th ed). APS Press, St. Paul, MN.

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

Hartwig, E.E. 1973. Varietal development. p. 187-210. In Soybeans: improvement, production and uses. American Society of Agronomy, Madison, WI.

Hartwig, E. E., and Kilen, T. C. 1991. Yield and composition of soybean seed from parents with different protein, similar yield. Crop Sci. 31:290-292.

Hartwig, E. E., Kuo, T. M., and Kenty, M. M. 1997. Seed protein and its relationship to soluble sugars in soybean. Crop Sci. 37:770-773.

Hashmi, R., Bond, J., Schmidt, M., and Klein, J. 2005. A temperature-controlled water bath method for evaluating soybean reaction to sudden death syndrome (SDS). Plant Health Prog. Online. <http://www.plantmanagementnetwork.org/pub/php/research/2005/sds/>.

Henriksen, B., and Elen, O. 2005. Natural *Fusarium* grain infection level in wheat, barley and oat after early application of fungicides and herbicides. J. Phytopath. 153:214-220.

Hershman, D., Hendrix, J., Stuckey, R., Bachi, P., and Henson, G. 1990. Influence of planting date and cultivar on soybean sudden death syndrome in Kentucky. Plant Dis. 74:761-766.

Hirrel, M. 1983. Sudden death syndrome of soybean: A disease of unknown etiology. Phytopathology. 73:42-54.

Hnetkovsky, N., Chang, S., Doubler, T., Gibson, P., and Lightfott, D. 1996. Genetic mapping of loci underlying field resistance to soybean sudden death syndrome (SDS). Crop Sci. 36:393-400.

Hurburgh, C., Jr., Brumm, T., Guinn, J., and Hartwig, R. 1990. Protein and oil patterns in U.S. and world soybean markets. J. Am. Oil Chem. Soc. 67:966-973.

Hwang, E. Y., Song, Q., Jia, G., Specht, J. E., Hyten, D. L., Costa, J., and Cregan, P. B. 2014. A genome-wide association study of seed protein and oil content in soybean. BMC Genomics. 15:1.

Hymowitz, T., Collins, F. I., Panczner, J., and Walker, W. M. 1972. Relationship between the content of oil, protein, and sugar in soybean seed. Agron. J. 64:613-616.

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

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

Illinois Soybean Association (ISA). 2013. Looking beyond soybean yields: Meeting customer needs with high protein and oil. Available at <http://www.ilsoy.org/sites/default/files/documents/1746.pdf>. Accessed March 3, 2017.

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

Jardine, D. J. and Todd, T. C. 2001. Soybean cyst nematode. L-786. Kansas State University Agricultural Experiment Station and Cooperative Extension Service.

Ji, J., Scott, M., and Bhattacharyya, M. 2006. Light is essential for degradation of Ribulose-1, 5-Bisphosphate Carboxylase-Oxygenase large subunit during sudden death syndrome development in soybean. Plant Biol. 8:597-605.

Jin, H., Hartman, G., Nickell, C., and Widholm, J. 1996. Characterization and purification of a phytotoxin produced by *Fusarium solani*, the causal agent of soybean sudden death syndrome. Phytopathology. 86:277-282.

Jun, T., Van, K., Kim, M. Y., Lee, S., and Walker, D. R. 2008. Association analysis using SSR markers to find QTL for seed protein content in soybean. Euphytica. 162:179-191.

Kabelka, E. A., Carlson, S. R., and Diers, B. W. 2005. Localization of Two Loci that Confer Resistance to Soybean Cyst Nematode from PI 468916. Crop Sci. 45:2473-2481.

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

Kao, C., and Zeng, Z. 1997. General formulas for obtaining the MLEs and the asymptotic variance- covariance matrix in mapping quantitative trait loci when using the EM algorithm. Biometrics. 53:653-665.

Kazi, S., Shultz, J., Afzal, J., Johnson, J., Njiti, V., and Lightfoot, D. A. 2008. Separate loci underlie resistance to root infection and leaf scorch during soybean sudden death syndrome. Theor. Appl. Genet. 116:967-977.

Kim, M., and Diers, B. W. 2013. Fine mapping of the SCN resistance QTL and from PI 468916. Crop Sci. 53:775-785.

Kim, M., Hyten, D. L., Bent, A. F., and Diers, B. W. 2010. Fine mapping of the SCN resistance locus from PI 88788. The Plant Genome. 3:81-89.

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

Koenning, S., and Barker, K. 1995. Soybean photosynthesis and yield as influenced by *Heterodera glycines*, soil type and irrigation. J. Nematol. 27:51.

Koenning, S. R., and Wrather, J. A. 2010. Suppression of soybean yield potential in the continental United States from plant diseases from 2006 to 2009. Plant Health Prog. Online. doi:10.1094/PHP-2010-1122-01-RS.

Kolander, T., Bienapfl, J., Kurle, J., and Malvick, D. 2012. Symptomatic and asymptomatic host range of *Fusarium virguliforme*, the causal agent of soybean sudden death syndrome. Plant Dis. 96:1148-1153.

Kopisch-Obuch, F., McBroom, R., and Diers, B. 2005. Association between soybean cyst nematode resistance loci and yield in soybean. Crop Sci. 45:956-965.

Krueger, K., Goggi, A. S., Mallarino, A. P., and Mullen, R. E. 2013. Phosphorus and potassium fertilization effects on soybean seed quality and composition. Crop Sci. 53:602-610.

Lee, S. H., Bailey, M. A., Mian, M. A. R., Carter, T. E., J., Shipe, E. R., Ashley, D. A., Parrott, W. A., Hussey, R. S., and Boerma, H. R. 1996. RFLP loci associated with soybean seed protein and oil content across populations and locations. Theor. Appl. Genet. 93:649-657.

Leffel, R. C. 1988. High protein lines and chemical constituent pricing in soybean. J. Prod. Agric. 1:111-115.

Leffel, R. C., and Rhodes, W. K. 1993. Agronomic performance and economic value of high-seed-protein soybean. J. Prod. Agric. 6:365-368.

Li, S., Hartman, G., and Chen, Y. 2009. Evaluation of aggressiveness of *Fusarium virguliforme* isolates that cause soybean sudden death syndrome. J. Plant Pathol. 91:77-86.

Li, W., Sun, D., Du, Y., Chen, Q., Zhang, Z., Qiu, L., and Sun, G. 2007. Quantitative trait loci underlying the development of seed composition in soybean (*Glycine max* L. Merr.). Genome. 50:1067-1077.

Lightfoot, D., and K. Meksem (Inventors). 2002. Isolated polynucleotides and polypeptides relating to loci underlying resistance to soybean cyst nematode and soybean sudden death syndrome and methods employing same. U.S. Pat. Appl. Publ. No. 2002144310.

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

Liu, K. 1997. Chemistry and nutritional value of soybean components. p. 25-113. In Soybeans: Chemistry, technology, and utilization. Springer US.

Liu, S., Kandoth, P. K., Warren, S. D., Yeckel, G., Heinz, R., Alden, J., Yang, C., Jamai, A., El-Mellouki, T., Juvale, P. S., Hill, J., Baum, T. J., Cianzio, S., Whitham, S. A., Korkin, D., Mitchum, M. G., and Meksem, K. 2012. A soybean cyst nematode resistance gene points to a new mechanism of plant resistance to pathogens. Nature. 492:256-260.

Luckew, A., Leandro, L., Bhattacharyya, M., Nordman, D., Lightfoot, D., and Cianzio, S. 2013. Usefulness of 10 genomic regions in soybean associated with sudden death syndrome resistance. Theor. Appl. Genet. 126:2391-2403.

Matson, A. L., and Williams, L. F. 1965. Evidence of a fourth gene for resistance to the soybean cyst nematode. Crop Sci. 5:477-477.

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

McLean, K. S., and Lawrence, G. W. 1993. Interrelationship of *Heterodera glycines* and *Fusarium solani* in sudden death syndrome of soybean. J. Nematol. 25:434-439.

Meksem, K., Doubler, T., Chancharoenchai, K., Nijti, N., Chang, S., Arelli, A. R., Cregan, P. E., Gray, L. E., Gibson, P. T and Lightfoot, D. 1999. Clustering among loci underlying soybean resistance to Fusarium solani, SDS and SCN in near-isogenic lines. Theor. Appl. Genet. 99:1131-1142.

Mudge, J., Concibido, V., Denny, R., Young, N., and Orf, J. 1996. Genetic mapping of a yield depression locus near a major gene for soybean cyst nematode resistance. Soybean Genetics Newsletter. 23:175-178.

Mueller, D., Hartman, G., Nelson, R., and Pedersen, W. 2002. Evaluation of glycine max germ plasm for resistance to *Fusarium solani* f. sp. *glycines*. Plant Dis. 86:741-746.

Mueller, D., Nelson, R., Hartman, G., and Pedersen, W. 2003. Response of commercially developed soybean cultivars and the ancestral soybean lines to *Fusarium solani* f. sp. *glycines*. Plant Dis. 87:827-831.

Murphy, P. 2008. Soybean proteins. p. 229-269. In: Johnson L., White P., Galloway R. (eds). Soybeans: chemistry, production, processing, and utilization. AOCS Press, Urbana, IL.

Nakajima, T., Mitsueda, T., and Charchar, M. J. D. 1993. Occurrence of soybean sudden-death syndrome caused by *Fusarium solani* in Brazil. p. 79. Abstract. In 7th Int Fusarium Workshop, Pennsylvania State University, University Park.

Navi, S. S., and Yang, X. 2008. Foliar symptom expression in association with early infection and xylem colonization by *Fusarium virguliforme* (formerly *F. solani* f. sp. *glycines*), the causal agent of soybean sudden death syndrome. Plant Health Prog. Online. doi:10.1094/PHP-2008-0222-01-RS

Niblack, T., Arelli, P., Noel, G., Opperman, C., Orf, J., Schmitt, D., Shannon, J., and Tylka, G. 2002. A revised classification scheme for genetically diverse populations of *Heterodera glycines*. J. Nematol. 34:279-288.

Niblack, T., Colgrove, A., Colgrove, K., and Bond, J. 2008. Shift in virulence of soybean cyst nematode is associated with use of resistance from PI 88788. Plant Health Prog. Online. doi:10.1094/PHP-2008-0118-01-RS

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

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

Njiti, V., Meksem, K., Iqbal, M., Johnson, J., Kassem, M. A., Zobrist, K., Kilo, V., and Lightfoot, D. 2002. Common loci underlie field resistance to soybean sudden death syndrome in Forrest, Pyramid, Essex, and Douglas. Theor. Appl. Genet. 104:294-300.

Njiti, V., Shenaut, M., Suttner, R., Schmidt, M., and Gibson, P. 1996. Soybean response to sudden death syndrome: Inheritance influenced by cyst nematode resistance in Pyramid× Douglas progenies. Crop Sci. 36:1165-1170.

Noel, G., Riggs, R., and Wrather, J. 1992. History, distribution, and economics. biology and management of the soybean cyst nematode. 1-13.

O'Donnell, K., Sink, S., Scandiani, M. M., Luque, A., Colletto, A., Biasoli, M., Lenzi, L., Salas, G., Gonzalez, V, Ploper, D., Formento, N., Pioli, R., Aoki., T., Yang, X., and Sarver., B. 2010. Soybean sudden death syndrome species diversity within North and South America revealed by multilocus genotyping. Phytopathology. 100:58-71.

Panthee, D. R., Pantalone, V. R., West, D. R., Saxton, A. M., and Sams, C. E. 2005. Quantitative trait loci for seed protein and oil concentration, and seed size in soybean. Crop Sci. 45:2015-2022.

Phansak, P., Soonsuwon, W., Hyten, D. L., Song, Q., Cregan, P. B., Graef, G. L., and Specht, J. E. 2016. Multi-population selective genotyping to identify soybean (*Glycine max* (L.) Merr.) seed protein and oil QTLs. G3: Genes| Genomes| Genetics. g3-116.

Piper, E., and Boote, K. 1999. Temperature and cultivar effects on soybean seed oil and protein concentrations. J. Am. Oil Chem. Soc. 76:1233-1241.

Ploper, D. 1993. Síndrome de la muerte súbita: Nueva enfermedad de la soja en el noroeste argentine. Avance Agroindustrial. Año 13. 54:5-9.

Probst, A. H., and Judd, R. W. 1973. Origin, US history and development, and world distribution. Agronomy. 1973.

Pudake, R. N., Swaminathan, S., Sahu, B. B., Leandro, L. F., and Bhattacharyya, M. K. 2013. Investigation of the *Fusarium virguliforme fvtox1* mutants revealed that the *FvTox1* toxin is involved in foliar sudden death syndrome development in soybean. Curr. Genet. 59:107-117.

Rao-Arelli, A., Anand, S. C., and Wrather, A. J. 1992. Soybean resistance to soybean cyst nematode race 3 is conditioned by an additional dominant gene. Crop Sci. 32:862-864.

Recker, J. R., Burton, J. W., Cardinal, A., and Miranda, L. 2014. Genetic and phenotypic correlations of quantitative traits in two long-term randomly mated soybean populations. Crop Sci. 54:939-943.

Rotundo, J. L., and Westgate, M. E. 2009. Meta-analysis of environmental effects on soybean seed composition. Field Crops Res. 110:147-156.

Roy, K., Hershman, D., Rupe, J., and Abney, T. 1997. Sudden death syndrome of soybean. Plant Dis. 81:1100-1111.

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

Roy, K. W., Patel, M. V., and Baird, R. E. 2000. Colonization of *Heterodera glycines* cysts by *Fusarium solani* form A, the cause of sudden death syndrome, and other fusaria in soybean fields in the midwestern and southern United States. Phytoprot. 81:57-67.

Rupe, J. C. 1989. Frequency and pathogenicity of *Fusarium solani* recovered from soybeans with sudden death syndrome. Plant Dis. 73:581-584.

Rupe, J. C., Correll, J. C., Guerber, J. C., Becton, C. M., Gbur, Jr, E. E., Cummings, M. S., and Yount, P. A. 2001. Differentiation of the sudden death syndrome pathogen of soybean, *Fusarium solani* f. sp. glycines, from other isolates of *F. solani* based on cultural morphology, pathogenicity, and mitochondrial DNA restriction fragment length polymorphisms. Can. J. Bot. 79:829-835.

Rupe, J., Robbins, R., and Gbur, E. 1997. Effect of crop rotation on soil population densities of *Fusarium solani* and *Heterodera glycines* and on the development of sudden death syndrome of soybean. Crop Prot. 16:575-580.

Sanogo, S., and Yang, X. 2001. Relation of sand content, pH, and potassium and phosphorus nutrition to the development of sudden death syndrome in soybean. Can. J. Plant Pathol. 23:174-180.

Scherm, H., and Yang, X. 1996. Development of sudden death syndrome of soybean in relation to soil temperature and soil water matric potential. Phytopathology. 86:642-649.

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

Schmitt, D., and Shannon, G. 1992. Differentiating soybean responses to *Heterodera glycines* races. Crop Sci. 32:275-277.

Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D. L., Song, Q., Thelen, J. J., Cheng, J., and Xu, D. 2010. Genome sequence of the palaeopolyploid soybean. Nature. 463:178-183.

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

Shannon, J. G., Wilcox, J. R., and Probst, A. H. 1972. Estimated gains from selection for protein and yield in the F4 generation of six soybean populations. Crop Sci. 12:824-826.

Shimura, E., and Hanson, W. D. 1970. Covariance analysis involving energy production and distribution among seed fractions by soybean genotypes, *Glycine max* (L. Merrill). Crop Sci. 10:242-246.

Sinclair, T. R., and de Wit, C. T. 1976. Analysis of the carbon and nitrogen limitations to soybean yield. Agron. J. 68:319-324.

Sinclair, T. R., and de Wit, C. T. 1975. Photosynthate and nitrogen requirements for seed production by various crops. Science. 189:565-567.

Slack, D., Riggs, R., and Hamblen, M. 1972. The effect of temperature and moisture on the survival of *Heterodera glycines* in the absence of a host. J. Nematol. 4:263-266.

Sneller, C. H. 1994. Pedigree analysis of elite soybean lines. Crop Sci. 34:1515-1522.

SoyBase and the Soybean Breeder's Toolbox. Soybean breeder's toolbox genetic map information, available at <http://www.soybase.org>. Accessed January 20, 2017.

Specht, J., Chase, K., Macrander, M., Graef, G., Chung, J., Markwell, J., Germann, M., Orf, J., and Lark, K. 2001. Soybean response to water. Crop Sci. 41:493-509.

Srour, A., Afzal, A. J., Blahut-Beatty, L., Hemmati, N., Simmonds, D. H., Li, W., Liu, M., Town, C., Sharma, H., Arelli, P. and Lightfoot, D. A. 2012. The receptor like kinase at Rhg1-a/Rfs2 caused pleiotropic resistance to sudden death syndrome and soybean cyst nematode as a transgene by altering signaling responses. BMC Genomics. 13:368.

Tewoldemedhin, Y. T., S.C. Lamprecht, J.J. Geldenhuys, and F.J. Kloppers. 2015. First report of soybean sudden death syndrome caused by *Fusarium virguliforme* in South Africa. Phytopathology. 105:1601-1611.

Thomas, J. M. G., Boote, K. J., Allen, L. H., Gallo-Meagher, M., and Davis, J. M. 2003. Elevated temperature and carbon dioxide effects on soybean seed composition and transcript abundance. Crop Sci. 43:1548-1557.

Triantaphyllou, A.C., and Hirschmann, H., 1962. Oogenesis and mode of reproduction in the soybean cyst nematode, *Heterodera glycines*. Nematologica. 7:235

Triwitayakorn, K., Njiti, V., Iqbal, M., Yaegashi, S., Town, C., and Lightfoot, D. 2005. Genomic analysis of a region encompassing *QRfs1* and *QRfs2*: Genes that underlie soybean resistance to sudden death syndrome. Genome. 48:125-138.

USDA, ARS, National Genetic Resources Program. Germplasm Resources Information Network- (GRIN). National Germplasm Resources Laboratory, Beltsville, Maryland, available: <http://www.ars-grin.gov/cgi-bin/npgs/html/obvalue.pl?51016>. Accessed January 13, 2017.

USDA-Economic Research Services. 2013. Soybeans and Oil Crops, available at <http://www.ers.usda.gov/briefing/Baseline/crops.htm>. Accessed June 3, 2013.

USDA-National Agricultural Statistics Service. 2016. Quick Stats, available at <http://www.nass.usda.gov/QuickStats/index2.jsp>. Accessed February 25, 2016.

Vaughn, J. N., Nelson, R. L., Song, Q., Cregan, P. B., and Li, Z. 2014. The Genetic Architecture of seed composition in soybean Is refined by genome-wide association scans across multiple populations. G3: Genes| Genomes| Genetics. 4:2283-2294.

Vick, C., Chong, S., Bond, J., and Russin, J. 2003. Response of soybean sudden death syndrome to subsoil tillage. Plant Dis. 87:629-632.

Vollmann, J., ElHadad, T., R. Gretzmacher, R., and Ruckenbauer, P. 1996. Seed protein content of soybean as affected by spatial variation in field experiments. Plant Breed. 115:501-507.

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

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

Warrington, C. V., Abdel-Haleem, H., Hyten, D. L., Cregan, P. B., Orf, J. H., Killam, A. S., Bajjalieh, N., Li, Z. and Boerma, H. R. 2015. QTL for seed protein and amino acids in the Benning× Danbaekkong soybean population. Theor. Appl. Genet. 128:839-850.

Wu, X., Blake, S., Sleper, D. A., Shannon, J. G., Cregan, P., and Nguyen, H. T. 2009. QTL, additive and epistatic effects for SCN resistance in PI 437654. Theor. Appl. Genet. 118:1093-1105.

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

Wehrmann, V. K., Fehr, W. R., Cianzio, S. R., and Cavins, J. F. 1987. Transfer of high seed protein to high-yielding soybean cultivars. Crop Sci. 27:927-931.

Wen, Z., Tan, R., Yuan, J., Bales, C., Du, W., Zhang, S., Chilvers, M., Schmitdt, C., Song, Q., Cregan, P. and Wang, D. 2014. Genome-wide association mapping of quantitative resistance to sudden death syndrome in soybean. BMC genomics. 15:809.

Wilcox, J. R. 1998. Increasing seed protein in soybean with eight cycles of recurrent selection. Crop Sci. 38:1536-1540.

Wilcox, J. R., and Cavins, J. F. 1995. Backcrossing high seed protein to a soybean cultivar. Crop Sci. 35:1036-1041.

Wilcox, J. R., and Guodong, Z. 1997. Relationships between seed yield and seed protein in determinate and indeterminate soybean populations. Crop Sci. 37:361-364.

Wilcox, J. R., and Shibles, R. M. 2001. Interrelationships among seed quality attributes in soybean. Crop Sci. 41:11-14.

Wilson, R.F. 2004. Seed composition. p. 621-677. In Boerma, H.R., and Specht, J.E. (ed.) Soybeans: improvement, production, and uses. 3rd ed. Agron. Monogr. 16. ASA, CSSA, and SSSA, Madison, Wi.

Wilson, R.F. 2008. Soybean: Market driven research needs. p.3-14. In Stacey, G. (ed.) Genetics and genomics of soybean. 2nd ed. Springer, New York, NY.

Wolf, R. B., Cavins, J. F., Kleiman, R., and Black, L. T. 1982. Effect of temperature on soybean seed constituents: Oil, protein, moisture, fatty acids, amino acids, and sugars. J. Am. Oil Chem. Soc. 59:230-232.

Wrather, J. A., and Koenning, S. R. 2006. Estimates of disease effects on soybean yields in the United States 2003 to 2005. J. Nematol. 38:173-180.

Wrather, J., Kendig, S., Anand, S., Niblack, T., and Smith, G. 1995. Effects of tillage, cultivar, and planting date on percentage of soybean leaves with symptoms of sudden death syndrome. Plant Dis. 79:560-562.

Wrather, J., and Koenning, S. 2009. Effects of diseases on soybean yields in the United States 1996 to 2007. Plant Health Prog. Online. doi:10.1094/PHP-2009-0401-01-RS

Xing, L., and Westphal, A. 2006. Interaction of *Fusarium solani* f. sp. *glycines* and *Heterodera glycines* in sudden death syndrome of soybean. Phytopathology. 96:763-770.

Yaklich, R. W., Helm, R. M., Cockrell, G., and Herman, E. M. 1999. Analysis of the distribution of the major soybean seed allergens in a core collection of accessions. Crop Sci. 39:1444-1447.

Yaklich, R. W., Vinyard, B., Camp, M., and Douglass, S. 2002. Analysis of seed protein and oil from soybean northern and southern region uniform tests. Crop Sci. 42:1504-1515.

Young, L. D. 1999. Soybeans resistant to *Heterodera glycines* populations attacking Hartwig cultivar. J. Nematol. 31:583.

CHAPTER 2: IMPACT OF SEED PROTEIN ALLELES FROM THREE SOYBEAN SOURCES ON SEED COMPOSITION AND AGRONOMIC TRAITS*

Abstract

Soybean [*Glycine max* (L.) Merrill] meal is one of the most important plant-based protein sources in the world. Developing cultivars high in seed protein concentration and seed yield is a difficult task because the traits have an inverse relationship. Over two decades ago, a protein quantitative trait loci (QTL) was mapped on chromosome (chr) 20, and this QTL has been mapped to the same position in several studies and given the confirmed QTL designation cqSeed protein-003. In addition, the *wp* allele on chr 2, which confers pink flower color, has also been associated with increased protein concentration. The objective of our study was to evaluate the effect of cqSeed protein-003 and the *wp* locus on seed composition and agronomic traits in elite soybean backgrounds adapted to the Midwestern USA. Segregating populations of isogenic lines were developed to test the *wp* allele and the chr 20 high protein QTL alleles from Danbaekkong (PI619083) and PI468916 at cqSeed protein-003. An increase in protein concentration and decrease in yield were generally coupled with the high protein alleles at cqSeed protein-003 across populations, whereas the effects of *wp* on protein concentration and yield were variable. These results not only demonstrate the difficulty in developing cultivars with increased protein and yield but also provide information for breeding programs seeking to improve seed composition and agronomic traits simultaneously.

*Submitted to Theoretical and Applied Genetics. Brzostowski, L.F. et al. 2017. Theor. Appl. Genet.

Introduction

Soybean is grown as a source of protein and oil, and soybean seed averages approximately 350 g kg^{-1} protein (130 g kg^{-1} moisture basis). The seed contains a good balance of the amino acids necessary to meet the dietary requirements of swine and poultry (Liu, 1997; Kerley and Allee, 2003; Cromwell, 2012), which makes it an exceptional source of protein meal for livestock and a leading source of plant-based protein in the world (Wilson, 2008; Cromwell, 2012). It is important for breeders to continue to develop soybean cultivars that maintain and improve current protein levels, so soybean will continue its prominence as a livestock feed.

There is considerable range in seed protein concentration in soybean germplasm accessions. In the USDA Soybean Germplasm Collection, there are accessions with protein concentrations, on a 130 g kg^{-1} moisture basis, as low as 276 g kg^{-1} and as high as 504 g kg^{-1} (USDA, 2017). Additionally, protein concentration is a highly heritable trait with reported heritabilities of up to 0.99 (Brummer et al., 1997; Chung et al., 2003; Eskandari et al., 2013).

Although a high heritability and a substantive range in genotypic values should make increasing seed protein concentration an obtainable objective for breeding programs, complex relationships between protein, oil, and yield have made it difficult to efficiently combine high values for each of these three desirable traits into a single cultivar. The negative correlations between protein and oil concentration and protein and yield have been well established (Hartwig and Kilen, 1991; Wilcox and Cavins, 1995; Sebolt et al., 2000; Wilcox and Shibles, 2001; Chung et al., 2003; Wilson, 2004; Eskandari, et al., 2013; Bandillo et al., 2015). The leading hypothesis for the negative correlations is the physiological relationship between nitrogen and carbon supply. Both nitrogen and carbon supply play a role in yield and seed composition and are affected by nitrogen accumulation, partitioning, and whole plant remobilization (Sinclair and deWit, 1975). Protein, oil, and yield rely on the same limited carbon energy supply, and each component has a different energy requirement (Hanson, et al., 1961; Shimura and Hanson, 1970; Chung et al., 2003). On a genetic basis, hypotheses for the negative genetic correlations between protein and yield and protein and oil include but are not limited to pleiotropic effects or linkage drag (Chung et al., 2003; Nichols et al., 2006; Bandillo et al., 2015).

While the negative correlation between yield and protein concentration is strong, it is weaker than that between protein and oil (Chung et al., 2003). There is evidence that the

relationship between yield and protein can be modulated. Individual lines and groups of lines with elevated protein and yield have been identified in studies in which a high protein phenotype present in a low yielding parent has been backcrossed into a low protein, high yielding parent (Wehrmann et al., 1987; Wilcox and Cavins, 1995). In addition, recurrent selection and restricted index selection have been successfully used in developing high protein, high yielding lines (Brim and Burton, 1979; Holbrook et al., 1989). Following 26 generations of random mating to reduce linkage disequilibrium, Recker et al. (2014) observed no significant genetic correlations between yield and protein. However, a significant negative correlation between oil and protein was still observed, which provides further evidence of a pleiotropic relationship between the two traits. The aforementioned studies suggest there can be success in increasing yield and protein simultaneously when the appropriate breeding strategy is implemented. Evaluation and characterization of QTL associated with protein concentration can provide valuable information to help determine the best breeding scheme to meet seed composition and yield objectives.

Quantitative trait loci (QTL) for protein concentration have been mapped to all soybean chromosomes (Soybase, 2017). In one of the first QTL mapping studies in soybean, Diers et al. (1992) mapped two seed protein QTL in a population derived from a cross between the *G. max* experimental line, A81-356022, and the *Glycine soja* Siebold & Zucc. plant introduction, PI468916. One protein QTL was mapped to chromosome (chr) 15 (formerly linkage group (LG) E), whereas the other mapped to chr 20 (formerly LG I). These QTL were confirmed based on guidelines set forth by the Soybean Genetics Committee (Soybase). The QTL on chr 15 was given the designation cqProt-001 (Fasoula et al., 2004), and the QTL on chr 20 was designated cqProt-003 (Nichols et al., 2006). These designations have since been updated on the Soybase website (Soybase, 2017) and are now listed as cqSeed protein-001 and cqSeed protein-003. Protein QTL have been mapped from several sources to the same genomic regions on chr 15 (Lee et al., 1996; Brummer et al., 1997; Fasoula et al., 2004; Kim et al., 2015; Phansak et al., 2016) and chr 20 (Brummer et al., 1997; Sebolt et al., 2000; Chung et al., 2003; Wang et al., 2014; Warrington et al., 2015; Phansak et. al, 2016) suggesting these loci may have several alleles or the same alleles may be in several accessions or alternatively, there could be multiple closely linked QTL in these intervals. Follow-up studies have sought to refine the locations of the chr 15 and chr 20 QTL using advanced genetics techniques (Bolon et al., 2010; Hwang et al., 2014, Vaughn et al., 2014; Kim et al., 2015; Bandillo et al., 2015).

The Korean cultivar Danbaekkong (PI619083) contains a high protein allele at the chr 20 QTL (Harris, 2001; Yates, 2006; Warrington et al., 2015). Danbaekkong is a late maturity group (MG) IV soyfood cultivar (Kim et al, 1996). Although PI468916 and Danbaekkong have protein QTL that map to the same region on chr 20, it is unknown whether their alleles are the same or different. We will herein refer to the high protein QTL allele from Danbaekkong as CHR20-D and the high protein QTL allele from PI468916 as CHR20-PI.

CHR20-PI has been evaluated across northern U.S. soybean backgrounds and was found to be associated with greater protein, reduced yield, reduced oil, smaller seeds, taller plants, and/or earlier maturity (Sebolt et al., 2000; Nichols et al., 2006). Evaluation of CHR20-D in southern U.S. backgrounds and locations has shown an inconsistent association with yield, and it has been suggested that the Danbaekkong high protein allele could be successfully used to develop lines with high protein and yield (Harris, 2001; Yates, 2006). A recently released MGIII cultivar containing CHR20-D was demonstrated to have increased protein concentration and no yield loss compared to the checks (Mian et al., 2017). However, CHR20-D has not been directly evaluated in northern germplasm, and thus there is a need to determine whether its effect on protein and other traits is similar to what was observed by the high protein allele for CHR20-PI.

The recessive *wp* allele, which confers pink flower color, was mapped to chr 2 (formerly LG D1b), and this allele was found to be associated with increased seed protein concentration (Stephens and Nickell, 1992; Stephens et al. 1993). Stephens et al. (1993) also showed that the *wp* allele was associated with larger seeds and decreased seed oil concentration (Stephens et al. 1993). Hegstad et al. (2000) observed lines containing the *wp* allele in two populations had increased protein concentration, decreased oil concentration, later maturity, and increased plant height. Additionally, significant yield reductions associated with *wp* were observed in one population. Zabala and Vodkin (2005) determined that the pink flower color caused by the *wp* allele was the result of the insertion of a transposable element in the flavanone 3-hydroxylase gene 1. To date, the *wp* allele has not been tested in a background other than the one in which it was first discovered.

Before a protein-increasing QTL or gene can be widely used in breeding programs, it is important to analyze its effect, not only on protein concentration, but also agronomic traits, especially yield, in various high-yield genetic backgrounds. The objective of this study is to test

the effect of CHR20-D, CHR20-PI, and *wp* on protein concentration and other agronomic traits in multiple genetic backgrounds.

Materials and Methods

Plant Material

Population Development CHR20-D

Two populations of isogenic lines segregating for CHR20-D were developed. The donor parent Danbaekkong was mated to the recurrent parents 'Dwight,' a late MG II cultivar (Nickell et al., 1998), and LD02-5025, a late MG II elite breeding line (Cary and Diers, 2007). An F_2 plant that was homozygous for CHR20-D was selected from each mating and backcrossed to the respective recurrent parent using simple sequence markers (SSR) linked to the chr 20 QTL to facilitate the introgression without the need to analyze the seed protein contents of backcross progeny. An additional generation of backcrossing was conducted to reach the backcross-two F_1 (BC_2F_1) generation. After each generation of backcrossing, the presence of the CHR20-D allele was verified using several SSR markers linked to the QTL including Satt614, Satt239, and Satt354 (Nichols et al., 2006). Heterozygous BC_2F_1 plants were selfed to produce BC_2F_2 seed. Plants in this selfed generation were genotyped with markers linked to the QTL to identify plants homozygous for (1) the high protein allele (i.e., CHR20-D) from the donor and (2) the corresponding low protein allele from the recurrent parent. Any lines exhibiting a recombination between the SSR donor and recurrent parent markers were discarded. Two populations of BC_2F_2-derived lines, one for each recurrent parent, plus their respective recurrent parents and check cultivars were grown in the field in 2013 and 2014.

Population Development CHR20-PI and wp

Four populations were developed from four separate backcrosses (BC_4) in which one of four different Illinois-adapted genotypes were used as a recurrent parent. These parents included the two maturity group II cultivars Dwight (Nickell et al., 1998) and Loda (Nickell et al., 2001), and the two maturity group IV experimental lines LS93-0375 (Schmidt and Klein, 1993) and

C1981 (Nowling, 2001). The donor parent possessing the high protein CHR20-PI allele originated from a BC_3F_4 population (A81-356022 (4) x PI 468916) described by Sebolt et al. (2000). The donor parent for the *wp* allele was a F_4-derived line from the cross of two parents with pink flowers, LN89-5320 x LN89-5322 (Stephens and Nickell, 1992; Stephens et al., 1993).

The high protein QTL alleles in the two donor parents were introgressed into each of the four genetic backgrounds in the following manner. Presence of CHR20-PI was verified in BC_nF_1 plants with the SSR markers Satt239 and Satt496 (Nichols et al., 2006). Lines with a recombination between the two markers were discarded, and selected BC_nF_1 plants containing CHR20-PI were then mated to the recurrent parents. The presence of the *wp* allele was verified by performing progeny tests (ie. progenies fixed for green hypocotyl color inferred parent was homozygous for *wp*, progenies fixed for purple hypocotyl color inferred parent was homozygous for *Wp*, etc.) with the BC_nF_2 seed and occurred after the next backcross (BC_{n+1}) had taken place. Progeny tests from the previous generation were used to identify the $BC_{n+1}F_1$ seed to be genotyped with molecular markers to verify the presence of the CHR20-PI allele from PI468916. BC_4F_1 plants predicted to be heterozygous for alleles at both QTLs within each background were selfed, and single-seed descent was performed to produce BC_4F_4 seed. BC_4F_4 plants homozygous in both QTLs were selected and selfed to form populations of BC_4F_4-derived lines. Molecular markers linked to CHR20-PI and progeny tests for the *wp* locus were used to assess the genotype of the lines, and lines with recombination between markers in the region were discarded.

Field Evaluation

Environments and Check Cultivars CHR20-D Populations

In 2013 and 2014, populations of BC_2F_2–derived lines were evaluated at the Crop Sciences Research and Education Center in Urbana, IL and in a grower's field near Pontiac, IL. Planting dates were as follows: Pontiac, IL 2013, May 14; Urbana, IL 2013, May 15; Pontiac, IL 2014, May 7; Urbana, IL 2014, May 21. The check cultivar was IA2102 (Crochet and Hughes, 2011) for both populations. There were 39 lines in the LD00-5025 population (18 homozygous for CHR20-D at markers linked to the chr 20 QTL and 21 homozygous for the low protein allele at markers linked to the chr 20 QTL) and 47 lines in the Dwight population (24 homozygous for

CHR20-D at markers linked to the QTL, 23 homozygous for the low protein allele at markers linked to the chr 20 QTL).

Environments and Check Cultivars CHR20-PI and wp Populations

MG II BC_4F_4 Populations

Maturity group II BC_4F_4 populations were grown at the Northern Illinois Agronomy Research Center in DeKalb, IL in 2008, the Crop Sciences Research and Education Center in Urbana, IL in 2007 and 2008, a Mead, NE rainfed (Rf) location in 2007, and a Mead, NE irrigated (Ir) location in 2007 for a total of five environments. Planting dates were as follows: Urbana, IL 2007, May 16; Mead Rf, NE 2007 and Mead Ir, NE 2007, May 17; DeKalb, IL 2008, May 20; Urbana, IL 2008, May 28. Check cultivars were LD02-4485 (Abney and Crochet, 2006) and IA2068 (Abney and Crochet, 2003) in the Loda backcross population whereas the Dwight backcross population included only LD02-4485. The respective recurrent parent for each population was also included in the trials. There were 65 lines in the Loda population and 71 lines in the Dwight population.

MGIV BC_4F_4 Populations

Maturity group IV BC_4F_4 populations were planted at the Crop Sciences Research and Education Center in Urbana, IL during 2007 and 2008, a Mead, NE Rf location in 2007, and a Mead, NE Ir location in 2007 for a total of four environments. Planting dates were the same as those previously mentioned for the MG II populations. LD00-3309 (Diers et al., 2006) was a check cultivar in both MG IV populations while the LS93-0375 population included the cultivar Macon (Nickell et al., 1996) as an additional check. The recurrent parent for each population was also included in the field evaluations. There were 75 lines in the LS93-0375 population and 49 lines in the C1981 population.

Field Evaluation and Phenotypic Measurements for All Populations

Populations were blocked separately, and the lines plus the recurrent parents and check cultivars were arranged in a randomized complete block design. The CHR20-D populations were

grown in non-replicated tests, and the CHR20-PI populations were replicated twice. All populations were planted in two-row plots, 3.6 m long using a four-row ALMACO plot planter (ALMACO Nevada, Iowa). Row spacing was 0.76-m, and seeding rate was ~27 seeds per meter. All environments were rain-fed with the exception of Mead, NE (Ir). Plots were rated for maturity date, plant height, and lodging. Plant height was measured in cm as the distance between the soil surface and the top node on the main stem. Maturity was the date when 95% of the pods reached mature color (R8 described by Fehr et al., 1971) with September 1 recorded as 901. Lodging was rated on a scale of 1 and 5, with 1 equaling all plants erect and 5 equaling all plants prostrate. Seed yield was measured at maturity using a plot combine, adjusted to 130 g kg^{-1} moisture, and reported as kg ha^{-1}. Additionally, a Perten DA 7250 NIR analyzer was used to determine protein and oil concentration on a 130 g kg^{-1} moisture basis for the CHR20-D populations (Perten Hagersten, Sweden). Seed protein and oil concentration analysis for the CHR20-PI and *wp* locus populations was performed at the USDA Northern Regional Research Center in Peoria, IL and also reported on a 130 g kg^{-1} moisture basis.

DNA Extraction and Genetic Marker Analysis for All Populations

Genomic DNA was isolated from young trifoliolate leaves by a modified CTAB method described by Keim et al. (1988) or a quick DNA extraction method described by Bell-Johnson et al. (1998). Polymorphic simple sequence repeat (SSR) markers were used to perform polymerase chain reactions according to Cregan and Quigley (1997). Amplification products were separated in 6% (w/v) non-denaturing polyacrylamide gels by electrophoresis (Wang et al., 2003).

Statistical Analysis for All Populations

All data were subjected to analysis of variance using SAS v9.4 (SAS Institute Inc., Cary, NC) PROC MIXED. Data were analyzed across and within locations and an environment was a year by location combination. Marker genotype and lines nested within marker genotype were considered to be fixed effects whereas replicate and environment were treated as a random effect. Degrees of freedom were calculated according to the Kenward-Roger method (Littell et al., 2006).

Results

CHR20-D

CHR20-D was evaluated in the LD02-5025 and Dwight backgrounds, and each population was evaluated for seed composition and agronomic traits at four environments in Illinois. For both backgrounds, marker alleles from Danbaekkong linked to CHR20-D were associated with a significant ($P<0.05$) increase in protein concentration, decreased oil concentration, and increased lodging score compared to the recurrent parent allele across environments (Table 2.1; Table 2.2). In addition, lines containing CHR20-D had a significant ($P<0.0001$) yield reduction across environments compared to lines containing the recurrent parent allele for both backgrounds (Table 2.1). This difference was -455 kg ha^{-1} in the LD02-5025 background and -363 kg ha^{-1} in the Dwight background, which represent a seed yield decrease associated with the introgression of the donor parent high protein allele. Maturity date was not significant ($P<0.05$) over environments in the LD02-5025 population, but was significant for the Dwight population with lines containing the Dwight allele maturing two days earlier than lines with the Danbaekkong allele (Table 2.2). Additionally, a significant marker genotype x environment interaction was observed for protein and oil concentration in both populations. The marker genotype x environment interactions for yield were non-significant.

For the LD02-5025 population, the lines containing the high protein QTL allele had increased average protein concentration and decreased oil concentration for each environment with the exception of Pontiac in 2013 (Table 2.1). These significant differences ranged from 25 to 31 g kg^{-1} for protein concentration and -10 to -14 g kg^{-1} for oil concentration. Within all four environments, lines with the high protein QTL allele on average yielded significantly ($P<0.05$) less than lines with the LD02-5025 allele, and this difference ranged from -273 to -558 kg ha^{-1}.

Similar trends were observed in the Dwight population within environments for protein concentration, oil concentration, and yield (Table 2.1). Lines with the high protein QTL allele had significantly increased average protein concentration and decreased oil concentration compared to lines with the Dwight allele in the Urbana 2013, Urbana 2014, and Pontiac 2014 environments. These significant differences ranged from a 19 to 28 g kg^{-1} increase in protein concentration and a coupled -7 to -14 g kg^{-1} decrease in oil concentration. In addition, lines with

CHR20-D yielded significantly less than those with the Dwight allele in all four environments. The observed difference ranged from -239 at Pontiac 2014 to -496 kg ha^{-1} at Urbana 2014.

CHR20-PI and wp

Four populations were developed via backcrossing to test the effect of CHR20-PI and *wp* on seed composition and agronomic traits. The Loda and Dwight populations were evaluated in five environments while the LS93-0375 and C1981 populations were evaluated in four environments. Across environments, a significant marker genotype x environment interaction was detected for protein within all backgrounds. Additional significant marker genotype x environment interactions were population-specific.

Within and across environments, CHR20-PI was associated with significantly increased protein concentration and decreased oil concentration compared to the recurrent parent allele for all four populations (Table 2.3). The magnitude of the effect was dependent upon genetic background and environment. Within and across environments, the effect of CHR20-PI on yield was variable, although that variability did not include an example of a significant yield increase. Across environments, lines containing CHR20-PI had significantly reduced yields in the Dwight and C1981 populations, but such lines in the Loda and LS93-0375 populations did not exhibit significant yield depression. Within each population, CHR20-PI significantly decreased maturity date one to three days across environments (Table 2.2). A significant increase in plant height was also observed across environments in the Loda, Dwight, and LS93-0375 populations with plants containing the donor allele averaging 2.1 to 3.3 centimeters taller than those containing the recurrent allele. Significant associations were not observed for lodging in any of the four populations.

While CHR20-PI was consistently associated with an increase in protein concentration and a decrease in oil concentration, the *wp* allele had a non-significant effect on oil concentration and a variable effect on protein concentration across environments when lines homozygous for *wp* were compared to lines containing no high protein alleles (Table 2.4). Across environments, the *wp* allele also had a variable effect in terms of significance when lines homozygous for *wp* were compared to lines with no high protein alleles on yield, maturity date, plant height within the Loda, Dwight, LS93-0375, and C1981 backgrounds. When lines contained both the *wp* allele

and CHR20-PI, a significant increase in protein concentration was observed in all backgrounds in comparison to lines containing no high protein alleles; however, yield and oil concentration were significantly decreased.

Discussion

Although CHR20-PI has been studied for over two decades, detailed seed concentration and agronomic information on CHR20-D and the *wp* locus is more limited. Our study evaluated CHR20-PI, CHR20-D, and the *wp* alleles to determine whether these genes can be effectively used to improve seed composition in a breeding program targeted at improving the seed protein concentration in high-yield cultivar development.

For the most part, similar seed composition and yield trends were observed when the Danbaekkong high protein allele was introgressed into the Dwight and LD02-5025 backgrounds. This is not surprising because these two recurrent backgrounds not only have the same maturity but are also related with Dwight a parent of LD02-5025. In both populations, lines containing the Danbaekkong high protein allele had decreased yield across and within environments and also had increased protein and decreased oil across and within all environments with the exception of Pontiac 2013 (Table 2.1). We do not have a good explanation for the inconsistent Pontiac 2013 results, but it may have to do with the growing environment at this location during 2013, as seed composition is influenced by numerous environmental conditions such as temperature and moisture (Dornbos and Mullen, 1992; Gibson and Mullen, 1996; Specht et al., 2001; Carrera et al., 2009). The influence of the growing environment is supported by Pontiac 2013 having the lowest average protein concentrations of the four environments where the population was grown.

In previous studies using elite germplasm from the southern USA, CHR20-D was shown to have an inconsistent effect on seed yield in southern environments (Harris, 2001; Yates, 2006). This contrasts with the results from both Danbaekkong populations in our study where the Danbaekkong high protein allele was consistently associated with significantly decreased yield. Furthermore, we observed a significant yield decrease even when there was no significant increase in protein concentration. A number of explanations for the apparent discrepancy between our study and the previous studies include, but are not limited to, environmental influence, genetic background, and genetic linkage.

While CHR20-D was consistently associated with decreased yields, CHR20-PI had a more variable effect on yield. Additionally, CHR20-PI significantly increased protein and decreased oil within all environments and populations. This consistency was not observed for CHR20-D where a significant effect on seed composition was not seen in the Pontiac 2013 environment for both populations. Because we did not introgress high protein alleles of CHR20-D and CHR20-PI into the same genetic backgrounds and test them in the same environments, we cannot directly compare the effects of these two alleles. Therefore, we are unable to speculate on their allelic identity relationship based on this study. With that caveat noted, we did observe that CHR20-PI allele had a smaller effect on yield than did the CHR20-D allele.

For protein concentration, CHR20-PI was more consistent than the *wp* locus in increasing protein concentration across genetic backgrounds. In the C1981 population, lines containing the *wp* allele did not have a significant increase in protein concentration compared to lines containing no high protein alleles (Table 2.4). When the *wp* allele was stacked with CHR20-PI in this background, protein concentration was not numerically different than lines containing only CHR20-PI. Within the LS93-0375 background, the *wp* allele was ineffective in significantly increasing protein concentration on its own, but in combination with CHR20-PI, a significant increase in protein concentration in relation to lines with no high protein alleles was observed. Only in the Loda population were the *wp* allele and CHR20-PI numerically similar in their impact on protein concentration. Other than in the C1981 background, lines containing both the chr 20 and chr 2 protein increasing alleles had on average the greatest protein concentration compared to lines in the other three possible genotypic groups. CHR20-PI increased protein concentration, but also was associated with decreased oil across genetic backgrounds and environments (Table 2.3). CHR20-PI also was associated with decreased yield and increased plant height variably across environments and genetic backgrounds (Table 2.2 and 3). Stacking *wp* in combination with CHR20-PI generally produced results that would be expected if two-locus interaction was not significant (i.e., the two alleles at each locus interacted in an additive fashion) for all traits across environments (Table 2.4). The combination of the high protein alleles at the chr 20 and chr 2 loci increased protein concentration to the greatest extent; however, this combination also decreased seed yield to the greatest extent numerically across genetic backgrounds. The reliability of CHR20-PI for increasing protein concentration would make it a better candidate than *wp* for a forward breeding application. However, if yield is the

primary goal, neither allele would likely be a successful candidate in breeding program aimed at developing high yield cultivars through a traditional marker-assisted selection (MAS) breeding scheme.

Rapid improvements in genotyping and big data set analysis have led to recent protein and oil QTL mapping studies using diverse, large populations and with high density genetic markers (Hwang et al., 2014; Bandillo et al., 2015; Vaughn et al., 2015; Phansak et al., 2016; Qi et al., 2016). While additional seed composition QTL have been mapped in these studies, the chr 20 QTL region continues to be identified as having the largest effect on protein and oil concentration. Data from these studies can be used to better characterize and define the chr 20 QTL and ultimately clone it.

As more information is generated about genes that control seed composition, this information can not only be used to dissect the genetic architecture of composition and generate more efficient markers for MAS, but also to provide insight into the relationship between seed composition and yield. Even with rapid advances in QTL mapping technologies and methods, QTL confirmation and evaluation studies remain important so that mapped QTL can be effectively incorporated into a breeding program to improve seed composition traits.

Predictive modeling has shown promise to revolutionize plant breeding by improving genetic gain through a decrease in the length of breeding cycles and an increase in selection accuracy. Prediction accuracies over 0.60 have been reported for yield, protein, and oil, and it is assumed that these accuracies will further increase with improved statistical models and methods (Jarquin et al., 2014; Xavier et al., 2016; Jarquin et al., 2016). QTL mapping and evaluation studies can be important tools to aid breeders in selecting the most appropriate prediction model, making the model more robust, or assembling a strong training population. Overall, improved genomic selection techniques have potential to lead to the development of more high protein and high yield cultivars.

The development of cultivars with improved yield and protein concentration continues to be challenging due to the negative relationship between the two traits. The QTL evaluated in this study, and in other studies where protein and yield were both evaluated, provide genetic evidence for this negative correlation (Hegstad et al., 2000; Sebolt et al., 2000; Chung et al., 2003; Nichols et al., 2006). We cannot demonstrably document whether the impact on both protein and oil of the two alleles at the chr 20 and chr 2 QTLs that we studied arose from single-locus pleiotropy or

two-locus linkage. However, the inability of researchers to separate the effect of the QTL on both traits and the high energy cost of producing protein suggests that it is likely pleiotropy. The continued evaluation of QTL combined with advancements in genetic technologies could help us better understand the genetic relationships among seed components and lead to better strategies to develop cultivars with increased protein concentration and yield.

Tables

Table 2.1. The impact on seed yield, protein, and oil when the protein-increasing allele of the chr 20 QTL from Danbaekkong (CHR20-D) was introgressed into the LD02-5025 and Dwight backgrounds. Protein and oil concentrations are reported on a 130 g kg^{-1} moisture basis.

		Seed yield (kg ha^{-1})				Protein (g kg^{-1})				Oil (g kg^{-1})			
Genetic Background†	Environment‡	Donor allele§	Recurrent allele¶	Difference#	*P* value	Donor allele	Recurrent allele	Difference	*P* value	Donor allele	Recurrent allele	Difference	*P* value
LD02-5025	Pontiac 2013	4188	4650	-462	<.0001	338	342	-4	ns	182	181	1	ns
	Urbana 2013	3741	4299	-558	<.0001	374	343	31	<.0001	170	184	-14	<.0001
	Pontiac 2014	4591	5116	-525	<.0001	373	348	25	<.0001	166	176	-10	<.0001
	Urbana 2014	3889	4162	-273	0.0047	367	337	30	<.0001	169	181	-12	<.0001
	Across	4102	4557	-455	<.0001	363	343	20	<.0001	172	181	-9	<.0001
Dwight	Pontiac 2013	3598	3875	-277	0.001	350	350	0	ns	177	177	0	ns
	Urbana 2013	3339	3778	-439	<.0001	378	350	28	<.0001	171	185	-14	<.0001
	Pontiac 2014	4149	4645	-496	<.0001	369	346	23	<.0001	171	179	-8	<.0001
	Urbana 2014	3439	3678	-239	0.003	369	343	26	<.0001	172	181	-9	<.0001
	Across	3631	3994	-363	<.0001	366	347	19	<.0001	173	180	-7	<.0001

ns = non-significant.

†Recurrent parent of population.

‡Location and year.

§Mean of lines predicted to be homozygous for the high protein Danbaekkong allele at the chr 20 QTL based on the genetic markers Satt614, Satt239, and Satt354.

¶ Mean of lines predicted to be homozygous for the recurrent parent low protein allele at the chr 20 QTL based on the genetic markers Satt614, Satt239, and Satt354.

#Difference between the means of lines that were homozygous for the donor and recurrent parent allele classes.

Table 2.2. The impact on maturity, plant height, and plant lodging when the protein-increasing allele on chr 20 derived from Danbaekkong (CHR20-D) or derived from PI468916 (CHR20-PI) was introgressed into the listed genetic backgounds.

		Maturity date†				Plant Height (cm)‡				Lodging (1-5)§			
QTL allele¶	Genetic Background#	Donor allele††	Recurrent allele‡‡	Difference§§	*P* value	Donor allele	Recurrent allele	Difference	*P* value	Donor allele	Recurrent allele	Difference	*P* value
CHR20-D	LD02-5025	919	920	-1	ns	94.0	91.0	3.0	ns	2.7	2.2	0.5	<.0001
	Dwight	918	920	-2	0.0016	94.3	93.2	1.1	ns	2.2	2.0	0.2	0.018
CHR20-PI	Loda	916	917	-1	0.0262	72.2	70.1	2.1	0.0174	2.1	2.0	0.1	ns
	Dwight	917	918	-1	0.0005	77.8	75.5	2.3	0.0004	1.7	1.5	0.2	ns
	LS93-0375	925	926	-1	0.0044	92.4	89.1	3.3	<.0001	1.5	1.6	-0.1	ns
	C1981	928	931	-3	0.0058	106.4	104.4	2.0	ns	2.1	2.0	0.1	ns

ns = non-significant.

†Characterized as the calendar date when 95% of pods have reached mature color (R8; Fehr et al., 1971) with September 1 equivalent to 901.

‡Measured as the distance from the soil surface to the topmost node on the main stem.

§Lodging is visually rated on a 1 to 5 scale (i.e., 1=all plants erect and 5=all plants prostrate).

¶High protein allele name (that originated from one or the other donor parent - see caption).

Recurrent parent of population.

†† Mean of lines predicted to be homozygous for the high protein CHR20-D allele based on the genetic markers Satt614, Satt239, and Satt354 or the high protein CHR20-PI allele based on the genetic markers Satt239 and Satt496.

‡‡ Mean of lines predicted to be homozygous for the chr 20 low protein allele of the recurrent parent 20 based on the genetic markers Satt614, Satt239, and Satt354 or Satt239 and Satt496.

§§ Difference between the means of homozygous classes.

Table 2.3. The impact on seed yield, protein, and oil when the high protein QTL allele on chr 20 from PI468916 (CHR20-PI) was introgressed into the Loda, Dwight, LS93-0375, and C1981 backgrounds. Protein and oil concentrations are reported on a 130 g kg^{-1} moisture basis.

		Seed yield (kg ha^{-1})				Protein (g kg^{-1})				Oil (g kg^{-1})			
Genetic Background†	Environment‡	Donor allele§	Recurrent allele¶	Difference#	*P* value	Donor allele	Recurrent allele	Difference	*P* value	Donor allele	Recurrent allele	Difference	*P* value
Loda	Dekalb, IL 2008	3138	3188	-50	ns	356	344	12	<.0001	174	181	-7	<.0001
	Mead, NE (Ir) 2007	2819	3092	-273	0.0031	392	370	22	<.0001	165	176	-11	<.0001
	Mead, NE (Rf) 2007	2859	3115	-256	0.0036	383	361	22	<.0001	171	181	-10	<.0001
	Urbana, IL 2007	2510	2672	-162	0.0241	381	363	18	<.0001	170	180	-10	<.0001
	Urbana, IL 2008	2024	1953	71	ns	362	349	13	<.0001	176	182	-6	<.0001
	Across	2670	2804	-134	ns	375	357	18	<.0001	171	180	-9	0.0004
Dwight	Dekalb, IL 2008	3659	3941	-282	<.0001	354	342	12	<.0001	165	174	-9	<.0001
	Mead, NE (Ir) 2007	3542	4020	-478	<.0001	386	360	26	<.0001	153	166	-13	<.0001
	Mead, NE (Rf) 2007	3486	3895	-409	<.0001	384	358	26	<.0001	159	171	-12	<.0001
	Urbana, IL 2007	3233	3510	-277	<.0001	364	347	17	<.0001	162	172	-10	<.0001
	Urbana, IL 2008	2853	2999	-146	ns	347	333	14	<.0001	175	184	-9	<.0001
	Across	3354	3673	-319	0.0053	367	348	19	0.0019	163	173	-10	<.0001
LS93-0375	Mead, NE (Ir) 2007	4256	4513	-257	0.0003	398	373	24	<.0001	156	169	-13	<.0001
	Mead, NE (Rf) 2007	4133	4339	-206	0.0058	394	369	25	<.0001	162	175	-13	<.0001
	Urbana, IL 2007	3712	3861	-149	0.0483	386	365	21	<.0001	164	177	-13	<.0001
	Urbana, IL 2008	3240	3213	27	ns	374	358	16	<.0001	174	183	-9	<.0001
	Across	3835	3982	-147	ns	388	366	22	<.0001	164	176	-12	<.0001
C1981	Mead, NE (Ir) 2007	4003	4239	-236	ns	403	378	25	<.0001	155	167	-12	<.0001
	Mead, NE (Rf) 2007	3989	4370	-381	0.0010	404	374	30	<.0001	157	171	-14	<.0001
	Urbana, IL 2007	3424	3655	-231	0.0042	386	364	22	<.0001	163	177	-14	<.0001
	Urbana, IL 2008	3098	3330	-232	0.0008	375	360	15	<.0001	168	179	-11	<.0001
	Across	3629	3899	-270	0.0007	392	369	23	0.0005	161	174	-13	<.0001

ns = non-significant.

†Recurrent parent of population.

Table 2.3 (cont)

‡Location and year.

§ Mean of lines predicted to be homozygous for CHR20-PI based on the genetic markers Satt239 and Satt496.

¶ Mean of lines predicted to be homozygous for the recurrent parent allele at chr 20 based on the genetic markers Satt239 and Satt496.

Difference between the means of homozygous classes.

Table 2.4. Across environment means for seed yield, protein, and oil, and for maturity, lodging and height of the genotypic class arising from the introgression of the chr 20 low protein allele from the donor parent PI468916 and the introgression of the chr 2 *Wp* allele (purple flower/low protein) from LN89-5320 or LN89-5322 (Stephens and Nickell, 1992; Stephens et al., 1993) compared to the deviations from that mean for the other three genotypic classes. Of the four listed genotype classes for each genetic background, the first represents lines whose two-locus genotype matches that of the recurrent parent (low protein - purple flower/low protein), the next two represent lines with one or the other single allele introgressions, and the forth represents lines with a dual introgression of both the high protein allele at chr 20 from PI468916 (CHR20-PI) and the pink flower/high protein allele (*wp*). Protein and oil concentrations are reported on a 130 g kg^{-1} moisture basis.

Genetic				Seed Yield	Seed Protein	Seed Oil	Maturity date¶	Lodging#	Plant Height††
Background†	Locus‡		n§	(kg ha^{-1})	(g kg^{-1})	(g kg^{-1})		(1-5)	(cm)
	chr 20‡‡	*wp*§§							
Loda	*low*	*Wp*	18	3004	352	182	916	2.1	72
	low	*wp*	17	-442	13**	-5	3**	-0.1	-5***
	high	*Wp*	17	-159	16***	-8***	-1	0.1	2*
	high	*wp*	13	-518**	31***	-14***	2	-0.1	-3**
Dwight	*low*	*Wp*	18	3790	346	174	918	1.6	78
	low	*wp*	22	-284***	4*	-2	-1	-0.1	-6***
	high	*Wp*	12	-279***	17***	-10***	-2**	0.2	3**
	high	*wp*	19	-603***	24***	-12***	-2***	0	-3***
LS93-0375	*low*	*Wp*	17	4127	366	176	925	1.5	91
	low	*wp*	19	-307*	0	0	2*	0	-5*
	high	*Wp*	16	-131	18***	-10***	-1	0	3**
	high	*wp*	23	-426**	25***	-13***	0	0	0
C1981	*low*	*Wp*	12	4064	367	174	930	2.1	105
	low	*wp*	11	-309***	0	0	1	0	-1
	high	*Wp*	11	-254***	25***	-13***	-2	0	8**
	high	*wp*	15	-567***	25***	-13***	-3*	-0.2	-3

Table 2.4 (cont)

*, **, *** Significant at the 0.05, 0.01, and 0.001 probability levels, respectively.

†Recurrent parent of population.

‡Genotype of the genotypic class.

§ Number of lines in the genotypic class.

¶ Characterized as the calendar date when 95% of pods have reached mature color (R8; Fehr et al., 1971) with September 1 equivalent to 901.

Distance between the soil line and the top node on the main stem.

†† Lodging is visually rated on a 1 to 5 scale with 1=all plants erect and 5=all plants prostrate.

‡‡ Genetic state at the chr 20 locus. 'low' is homozygous for the low protein allele, 'high' is homozygous for CHR20-PI.

§§ Genetic state at the *wp* locus. '*Wp*' is homozygous for the purple flower/low protein allele, and '*wp*' is homozygous for the pink flower/high protein allele.

References

Abney, T.S. and W.D. Crochet. 2003. Uniform soybean tests northern states, 2003. USDA-ARS, West Lafayette, IN.

Abney, T.S. and W.D. Crochet. 2006. Uniform soybean tests northern states, 2006. USDA-ARS, West Lafayette, IN.

Bandillo, N., D. Jarquin, Q. Song, R. Nelson, P. Cregan, J. Specht, and A. Lorenz. 2015. A population structure and genome-wide association analysis on the USDA soybean germplasm collection. The Plant Genome. 8.

Bell-Johnson, B., G. Garvey, J. Johnson, D. Lightfoot, and K. Meksem. 1998. Biotechnology approaches to improving resistance to SCN and SDS: Methods for high throughput marker assisted selection. Soybean Genet. Newsl. 25:115-117.

Bolon, Y.T., B. Joseph, S.B. Cannon, M.A. Graham, B.W. Diers, A.D. Farmer, G.D. May, G.J. Muehlbauer, J.E. Specht, Z.J. Tu and N. Weeks. 2010. Complementary genetic and genomic approaches help characterize the linkage group I seed protein QTL in soybean. BMC Plant Bio. 10:41.

Brim, C.A., and J.W. Burton. 1979. Recurrent selection in soybeans II: selection for increased percent protein in Seeds1. Crop Sci. 19:494-498.

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

Carrera, C., M.J. Martínez, J. Dardanelli, and M. Balzarini. 2009. Water deficit effect on the relationship between temperature during the seed fill period and soybean seed oil and protein concentrations. Crop Sci. 49:990-998.

Cary, T.R., and B.W. Diers. 2007. Northern regional soybean cyst nematode tests. University of Illinois. Urbana, IL.

Chung, J., H.L. Babka, G.L. Graef, P.E. Staswick, D.J. Lee, P.B. Cregan, R.C. Shoemaker, and J.E. Specht. 2003. The seed protein, oil, and yield QTL on soybean linkage group I. Crop Sci. 43:1053-1067.

Cregan, P., and C. Quigley. 1997. Simple sequence repeat DNA marker analysis. p 173-185. *In* G. Caetano-Anolles, and P.M. Gresshoff, P.M. (eds.) DNA markers: Protocols, applications, and overviews. John Wiley and Sons, New York.

Crochet, W.D. and T. Hughes. 2011. Uniform soybean tests northern states, 2009. USDA-ARS, West Lafayette, IN.

Cromwell, G.L. 2012. Soybean meal: An exceptional protein source. Soybean Meal InfoCenter, Ankeny, IA: available at <http://www.soymeal.org/ReviewPapers/SBMExceptionalProteinSource.pdf/>. Accessed January 30, 2017.

Diers, B.W., P. Keim, W.R. Fehr, and R.C. Shoemaker. 1992. RFLP analysis of soybean seed protein and oil content. Theor. Appl. Genet. 83:608-612.

Diers, B.W., T.R. Cary, D.J. Thomas, and C.D. Nickell. 2006. Registration of 'LD00-3309' soybean. Crop Science. 46:1384.

Dornbos, D.L., and R.E. Mullen. 1992. Soybean seed protein and oil contents and fatty acid composition adjustments by drought and temperature. J. Am. Oil Chem. Soc. 69:228-231. doi:10.1007/BF02635891

Eskandari, M., E.R. Cober, and I. Rajcan. 2013. Genetic control of soybean seed oil: II. QTL and genes that increase oil concentration without decreasing protein or with increased seed yield. Theor. Appl. Genet. 126:1677-1687.

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

Fehr, W., C. Caviness, D. Burmood, and J. Pennington. 1971. Stage of development descriptions for soybeans, *Glycine Max* (L.) Merrill. Crop Sci. 11:929-931.

Gibson, L.R., and R.E. Mullen. 1996. Soybean seed composition under high day and night growth temperatures. J. Am. Oil Chem. Soc. 73:733-737.

Hanson, W.D., R.C. Leffel, and R.W. Howell. 1961. Genetic analysis of energy production in the Soybean. Crop Sci. 1:121-126.

Harris, D.K. 2001. Genetic improvement of soybean seed traits and resistance to bud blight and root-knot nematodes. M.S. thesis, Univ. of Georgia, Athens.

Hartwig, E.E., and T.C. Kilen. 1991. Yield and Composition of Soybean Seed from Parents with Different Protein, Similar Yield. Crop Sci. 31:290-292.

Hegstad, J.M., L.O. Vodkin, and C.D. Nickell. 2000. Genetic and agronomic evaluation of *wp-m* in soybean. Crop Sci. 40: 346-351.

Holbrook, C.C., J.W. Burton, and T.E. Carter. 1989. Evaluation of recurrent restricted index selection for increasing yield while holding seed protein constant in soybean. Crop Sci. 29:324-329.

Hwang, E.Y., Q. Song, G. Jia, J.E. Specht, D.L. Hyten, J. Costa, and P.B. Cregan. 2014. A genome-wide association study of seed protein and oil content in soybean. BMC Genomics. 15:1.

Jarquin, D., K. Kocak, L. Posadas, K. Hyma, J. Jedlicka, G. Graef, and A. Lorenz. 2014. Genotyping by sequencing for genomic prediction in a soybean breeding population. BMC genomics. 15:740.

Jarquin, D., J. Specht, J., and A. Lorenz. 2016. Prospects of Genomic Prediction in the USDA Soybean Germplasm Collection: Historical data creates robust models for enhancing Selection of Accessions. G3: Genes| Genomes| Genetics. 6:2329-2341.

Kerley, M.S. and G.L. Allee. 2003. Modifications in soybean seed composition to enhance animal feed use and value: moving from a dietary ingredient to a functional dietary componment. AgBioForum. 6:14-17.

Keim, P., R. Olson, and R. Shoemaker. 1988. A rapid protocol for isolating soybean DNA. Soybean Genet. Newsl. 12:150-152.

Kim, M., S. Schultz, R. Nelson, and B.W. Diers. 2015. Identification and fine mapping of a soybean seed protein QTL from PI 407788A on chromosome 15. Crop Sci. 56:219-225.

Kim, S.D., E.H. Hong, Y.H. Kim, S.H. Lee, Y.K. Seong, K.Y. Park, Y.H. Lee, E.H. Hwang, E.H. Park, H.S. Kim, Y.H. Ryu, R.K. Park, and Y.S. Kim. (1996) A new high protein and good seed quality soybean variety "Danbaegkong". (In Korean.) RDA J. Agric. Sci.-Upland and Industrial Crops. 38:228-232.

Lee, S.H., M.A. Bailey, R.M. Mian, T.E. Carter, E.R. Shipe, D.A. Ashley, W.A. Parrot, R.S. Hussey, and H.R. Boerma. 1996. RFLP loci associated with soybean seed protein and oil content across populations and locations. Theor. Appl. Genet. 93:649-657.

Littell, R.C., G.A. Milliken, W.W. Stroup, R.D. Wolfinger, and O. Schabenberger. 2006. SAS® for Mixed Models, Second Edition. Cary, NC: SAS Institute Inc.

Liu, K. 1997. Soybeans: Chemistry, technology, and utilization. Chapman and Hall, New York.

Mian, R., L. McHale, Z. Li, and A.E. Dorrance. 2017. Registration of 'Highpro1' Soybean with high protein and high yield developed from a North × South cross. J. Plant. Reg. 11:51-54.

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

Nickell, C.D., D.J. Thomas, T.R. Cary, and D. Heavner. 1996. Registration of 'Macon' soybean. Crop Sci. 36:1410.

Nickell, C.D., G.R. Noel, T.R. Cary, D.J. Thomas, and R.A. Leitz. 1998. Registration of 'Dwight' soybean. Crop Sci. 38:1398.

Nickell, C.D., G.R. Noel, T.R. Cary, D.J. Thomas, and B.W. Diers. 2001. Registration of 'Loda' soybean. Crop Sci. 41:589-590.

Nowling, G.L. 2001. Uniform soybean tests northern states, 2001. USDA-ARS, West Lafayette, IN.

Phansak, P., W. Soonsuwon, D.L. Hyten, Q. Song, P.B. Cregan, G.L. Graef, and J.E. Specht. 2016. Multi-population selective genotyping to identify soybean (*Glycine max* (L.) Merr.) seed protein and oil QTLs. G3: Genes| Genomes| Genetics. 6:1635-1648.

Qi, Z., J. Pan, X. Han, H. Qi, D. Xin, W. Li, X. Mao, Z. Wang, H. Jiang, C. Liu, Z. Hu, G. Hu, R. Zhu, and Q. Chen. 2016. Identification of major QTLs and epistatic interactions for seed protein concentration in soybean under multiple environments based on a high-density map. Mol. Breed. 36:1-16.

Recker, J.R., J.W. Burton, A. Cardinal, and L. Miranda. 2014. Genetic and phenotypic correlations of quantitative traits in two long-term randomly mated soybean populations. Crop Sci. 54:939-943.

SAS Institute. 2016. The SAS system for Microsoft Windows. Release 9.4. SAS Inst., Cary, NC.

Schmidt, M.E. and J.H. Klein. 2002. Registration of 'LS93-0375' soybean. Crop Sci. 42:1375.

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

Shimura, E., and W.D. Hanson. 1970. Covariance analysis involving energy production and distribution among seed fractions by soybean genotypes, *Glycine max* (L.) Merrill. Crop Sci. 10:242-246.

Sinclair, T.R., and C.T. de Wit. 1975. Photosynthate and nitrogen requirements for seed production by various crops. Science. 189:565-567.

SoyBase and the Soybean Breeder's Toolbox. Soybean breeder's toolbox genetic map information, available at <http://www.soybase.org>. Accessed January 30, 2017.

Specht, J., Chase, K., Macrander, M., Graef, G., Chung, J., Markwell, J., Germann, M., Orf, J., and Lark, K. 2001. Soybean response to water. Crop Sci. 41:493-509.

Stephens, P.A. and C.D. Nickell. 1992. Inheritance of pink flower in soybean. Crop Science. 32:1131-1132.

Stephens, P. A., C.D. Nickell, and L.O. Vodkin. 1993. Pink flower color associated with increased protein and seed size in soybean. Crop Science. 33:1135-1137.

USDA-ARS, National Genetic Resources Program. Germplasm Resources Information Network- (GRIN). National Germplasm Resources Laboratory, Beltsville, Maryland, available: <https://npgsweb.ars-grin.gov/gringlobal> Accessed January 12, 2017.

Vaughn, J.N., R.L. Nelson, Q. Song, P.B. Cregan, and Z. Li. 2014. The genetic architecture of seed composition in soybean is refined by genome-wide association scans across multiple populations. G3: Genes| Genomes| Genetics. 4:2283-2294.

Wang, D., J. Shi, S. Carlson, P. Cregan, R. Ward, and B. Diers. 2003. A low-cost, high-throughput polyacrylamide gel electrophoresis system for genotyping with microsatellite DNA markers. Crop Sci. 43:1828-1832.

Wang, X., G.L. Jiang, M. Green, R.A. Scott, Q. Son, D.L. Hyten, and P.B. Cregan. 2014. Identification and validation of quantitative trait loci for seed yield, oil and protein contents in two recombinant inbred line populations of soybean. Mol. Genetics and Genomics. 289:935-949.

Warrington, C.V., H. Abdel-Haleem, D.L. Hyten, P.B. Cregan, J.H. Orf, A.S. Killam, N. Bajjalieh, Z. Li, and H.R. Boerma. 2015. QTL for seed protein and amino acids in the Benning× Danbaekkong soybean population. Theor. Appl. Genet. 128:839-850.

Wehrmann, V.K., W.R. Fehr, S.R. Cianzio, and J.F. Cavins. 1987. Transfer of high seed protein to high-yielding soybean cultivars. Crop Sci. 27:927-931.

Wilcox, J.R., and J.F. Cavins. 1995. Backcrossing high seed protein to a soybean cultivar. Crop Sci. 35:1036-1041.

Wilcox, J.R., and R.M. Shibles. 2001. Interrelationships among seed quality attributes in soybean. Crop Sci. 41:11-14.

Wilson, R.F. 2004. Seed composition. p 621-677. *In* Boerma, H.R., and Specht, J.E. (ed.) Soybeans: improvement, production, and uses. 3rd ed. Agron. Monogr. No. 16. ASA, CSSA, and SSSA, Madison, WI.

Wilson, R.F. 2008 Soybean: Market driven research needs. p 3-14. *In* Stacey, G. (ed.) Genetics and genomics of soybean. 2nd ed. Springer, New York, NY.

Yates, J.L. 2006. Use of diverse germplasm to improve peanut root-knot nematode resistance and seed protein content in soybean. Ph.D. diss., Univ. of Georgia, Athens.

Xavier, A., W.M. Muir, and K.M. Rainey. 2016. Assessing predictive properties of genome-wide selection in soybeans. G3: Genes| Genomes| Genetics. 6:2611-2616.

Zabala G., and L.O Vodkin. 2005. The *wp* mutation of *Glycine max* carries a gene-fragment-rich transposon of the CACTA superfamily. Plant Cell 17:2619-2632.

CHAPTER 3: AGRONOMIC EVALUATION OF A HIGH PROTEIN ALLELE FROM PI407788A ON CHROMOSOME 15 ACROSS TWO SOYBEAN BACKGROUNDS*

Abstract

Soybean protein [*Glycine max* (L.) Merrill] is a prominent plant-based protein source worldwide due to its high quality and relatively low cost. A major barrier to the development of high protein cultivars is the negative relationship between protein and yield. A large effect protein quantitative trait loci (QTL) has been mapped to the same location on chromosome (chr) 15 in several studies and given the designation, cqSeed protein-001. The objective of this study was to evaluate the effect of the high protein allele from PI407788A at the chr 15 locus on seed composition and agronomic traits. Segregating near-isogenic populations were formed by backcrossing the high protein allele into two elite soybean backgrounds, and these populations were planted at field environments in Illinois. Across backgrounds, the PI407788A allele significantly ($P<0.0001$) increased protein and decreased oil compared to the recurrent alleles but had a non-significant effect on yield. Information from this study will aid breeders in forming strategies to develop cultivars with increased seed composition and yield to meet the needs of a growing world population.

*Submitted to Crop Science. Brzostowski, L.F. and B.W. Diers. 2017. Crop Sci.

Introduction

Soybean protein is one of the highest quality plant-based protein sources in the world containing a good balance of amino acids to support the diet of non-ruminant livestock. Soybean meal (SBM), which is made out of crushed seed after oil is extracted from it, has the highest level of crude protein among plant-based protein sources and accounts for approximately 69% of the protein sources in animal feed worldwide (Wilson, 2008; Cromwell, 2012). The price of SBM is reduced if it is does not meet a 48% protein level, so it is important for breeders to develop soybean cultivars with elevated protein levels. Additionally, the value of soybean could increase if seed protein levels are improved.

Soybean seeds on average are 350 g kg^{-1} protein, 175 g kg^{-1} oil, 260 g kg^{-1} carbohydrates, 45 g kg^{-1} crude fiber, and 45 g kg^{-1} ash (130 g kg^{-1} moisture basis) (Hymowitz, 1972; Wilson, 2004). Complex relationships exist among these seed components and with seed yield. An increase in protein often leads to a decrease in oil, carbohydrates, and yield which is a major obstacle in the development of high protein cultivars that are profitable for producers (Leffel and Rhodes, 1993; Wilcox and Cavins, 1995; Hartwig et al., 1997; Wilcox and Shibles, 2001; Chung et al., 2003; Eskandari et al., 2013; Bandillo et al., 2015).

Several high protein cultivars have been released that have reported moderate yield loss compared to a check cultivar (Burton et al., 1999; Anand et al., 2004; Panthee and Pantalone, 2006; Carter et al., 2010; Chen et al., 2011). A limited number of cultivars have also been released that report no yield loss compared to a check cultivar, which suggests that it is possible to develop soybean with improved protein and yield using the appropriate combination of genetics, breeding strategy, and environments (Panthee and Pantalone, 2006; Mian et al., 2017); however, it should be noted that the aforementioned cultivars still had a significantly ($P<0.05$) decreased oil concentration in reference to check cultivars. While the mechanisms that control relationships between seed composition traits and yield are still unknown, hypotheses include genetic linkage and pleiotrophy (Chung et al., 2003; Nichols et al., 2006; Recker et al., 2014; Bandillo et al., 2015; Phansak et al., 2016).

Soybean protein concentration is inherited as a quantitative trait, and quantitative trait loci (QTL) for protein have been mapped to locations on all chromosomes (Soybase, 2017). Initial work in QTL mapping for seed composition traits was done using biparental populations.

Diers et al. (1992) identified two high protein QTL in a population derived from cross of an Iowa State University experimental line and the *Glycine soja* Sieb. and Zucc. plant introduction PI468916. One QTL was located on chromosome (chr) 20 (formerly linkage group (LG) I) and a second was located on chr 15 (LG E). These QTL have been confirmed based on guidelines set forth by the Soybean Genetics Committee (Soybase), and the QTL on chr 15 was given the designation cqSeed protein-001 (Fasoula et al., 2004) while the QTL on chr 20 was given the designation cqSeed protein-003 (Nichols et al., 2006). Large effect QTL have been mapped repeatedly to these two locations in other biparental populations and also more recently in diverse mapping panels (Lee et al., 1996; Brummer et al., 1997; Sebolt et al., 2000; Chung et al., 2003; Fasoula et al., 2004; Wang et al., 2014; Kim et al., 2015; Bandillo et al., 2015; Warrington et al., 2015; Phansak et al., 2016; Qi et al., 2016). Additionally, advances in genetic techniques and statistical analysis have led to subsequent studies that refined the location of the chr 15 and chr 20 QTL and identified candidate genes at the loci (Bolon et al., 2010; Hwang et al., 2014, Vaughn et al., 2014; Bandillo et al., 2015).

While the chr 15 QTL was confirmed prior to the chr 20 QTL, there has been much less research conducted on the chr 15 QTL especially in regard to its impact on agronomic traits. The chr 20 QTL has been consistently demonstrated to be associated with increases in seed protein and decreases in seed oil and yield (Sebolt et al., 2000; Chung et al., 2003; Nichols et al., 2006). Sebolt et al. (2000) evaluated the effect of the high protein allele on chr 20 from PI468916 across multiple genetic backgrounds, and the PI468916 allele was shown to be associated with increased protein, decreased yield and oil, smaller seeds, and an earlier maturity date. Similar observations were made in a follow-up study which also mapped the chr 20 QTL to a 3 cM interval (Nichols et al., 2006). Sebolt et al. (2000) did test for the chr 15 QTL, but no significant effects were detected, which was likely the result of the QTL being lost during backcrossing due to poor marker coverage in that study.

A QTL was mapped to the same region on chr 15 as cqSeed protein-001 in a BC_1F_5-derived population developed from a cross of the donor parent, LG00-13329, and the recurrent parent, 'Williams 82' (Kim et al., 2015). LG00-13329 had been previously derived from a cross of PI407788A, a high protein MG IV accession from Korea, to Williams 82. In the mapping population, the chr 15 QTL allele from PI407788A was associated with increased protein concentration and decreased oil concentration, and this QTL was fine mapped to a 535 kb

interval between BARCSOYSSR_15_0161 and BARCSOYSSR_15_0194. While this study provided useful information for marker-assisted selection (MAS) and efforts to clone the QTL, further studies of this refined genetic region are needed to determine whether this high protein allele is suitable for use in breeding programs focused on developing cultivars with high protein concentration and good agronomic performance.

It is important to evaluate a protein QTL across multiple genetic backgrounds and environments before it is incorporated into a breeding program due to the complex relationships between seed components and yield. While this step is necessary, there are currently a limited number of studies where this is a research objective, and this is especially true for the chr 15 protein QTL. Therefore, the objective of this study is to test the effect of the high protein allele from PI407788A located on chr 15 on seed composition and agronomic traits across two genetic backgrounds.

Materials and Methods

Population Development

Two populations of NILs segregating for the chr 15 QTL were developed through four backcrosses (BC_4). A moderate protein, maturity group II, elite experimental line AR09-192019 (Abney and Hughes, 2011) and a low protein, maturity group (MG) II, elite cultivar LD02-4485 (Abney and Crochet, 2006) were used as recurrent parents for each of the populations. LG05C-1782, was used as the donor parent for the chr 15 high protein QTL. LG05C-1782 is a BC_1 line that was developed as described by Kim et al. (2015) using Williams 82 as the recurrent parent and PI407788A as the donor of the high protein allele.

After each generation of backcrossing, the presence of the chr 15 high protein allele was confirmed using simple sequence repeat markers (SSR) linked to the QTL (Kim et al., 2015). Plants with the genetic region of interest were then used in the next generation of backcrossing until the BC_4F_1 generation was reached. BC_4F_1 plants heterozygous for the region of interest were identified and selfed to produce BC_4F_2 seed. BC_4F_2 plants homozygous for the donor and recurrent parent alleles of the chr 15 QTL region were identified with markers and grown in the greenhouse. Bulked seed from the individual plants were grown in the field as $BC_4F_{2:3}$ lines. Progeny from these lines were used for field evaluations in 2015 and 2016. There were 48 lines in the AR09-192019 background (24 homozygous for the PI407788A allele at chr 15 based on

markers linked to the QTL, 24 homozygous for the AR09-192019 allele at chr 15 based on markers linked to the QTL) and 84 lines in the LD02-4485 background (42 homozygous for the PI407788A allele at chr 15 based on markers linked to the QTL, 42 homozygous for the LD02-4485 allele at chr 15 based on markers linked to the QTL).

Field Evaluation and Phenotypic Measurements

In 2015 and 2016, a population of BC_4F_2-derived lines in the LD02-4485 background and a second population in the AR09-192019 background that were segregating for the chr 15 QTL were selected and placed into two tests based on maturity. Each test contained lines from both backgrounds, so the background effect on the QTL could be tested. The tests were planted at the Crop Sciences Research and Education center in Urbana, IL and in a grower's field near Pontiac, IL. Each test was grown in a separate randomized complete block design experiment that was replicated twice and included lines plus recurrent parents and checks. The tests were planted using an ALMACO plot planter (ALMACO Nevada, Iowa) at Pontiac 2015 on May 7, Urbana 2015 on May 14, Pontiac 2016 on May 20, and Urbana 2016 on May 23. The two-row plots were 3.6 m long with a row spacing of 0.76 m and a seeding rate of ~27 seeds per meter. Plant height, maturity data, and lodging score were collected on all plots prior to harvest. Plant height was the distance between the soil surface and the main stem apex in centimeters. Maturity date was the date on which 95% of pods reached their mature color (R8 described by Fehr et al., 1971); September 1 was recorded as 901. Lodging score was on a 1 to 5 scale, with 1 recorded as all plants within a row erect and 5 recorded as all plants within a row prostrate. At maturity, plots were harvested for yield using a plot combine. Yield data was adjusted to 130 g kg^{-1} moisture and reported as kg ha^{-1}. Protein and oil concentration on a 130 g kg^{-1} basis were collected on all lines using a Perten DA 7250 NIR analyzer (Perten Hagersten, Sweden).

DNA Extraction and Genetic Marker Analysis

Plants were genotyped with SSR markers linked to the QTL (Kim et al., 2015). Genomic DNA was extracted from young trifolioate leaf samples collected from individual plants via either a modified CTAB (Keim et al., 1988) or a quick DNA extraction (Bell-Johnson et al.,

1998) method. Polymerase chain reaction (PCR) was then performed using polymorphic genetic markers as described by Cregan and Quigley (1997). Electrophoresis was used to analyze the PCR products in 6% (w/v) nondenaturing polyacrylamide gels stained with ethidium bromide (Wang et al., 2003).

Statistical Analysis

All data were subjected to analysis of variance using SAS v9.4 (SAS Institute Inc., Cary, NC) PROC GLM. Data were analyzed across and within locations and genetic backgrounds. Data were also analyzed across the two tests, and an environment was a year by location combination. Marker genotype, test, and genetic background were fixed, and replication and environment were random.

Results

The near isogenic lines (NILs) were developed to be representative of the maturity group II germplasm grown in the Midwestern USA. LD02-4485 and AR09-192019 were originally selected because they were high yielding, similar in maturity, and differed in protein concentration based on previously reported agronomic data (Abney and Hughes, 2011). Significant differences ($P<0.0001$) between the genetic backgrounds were observed for all traits (Table 3.1).When evaluated across and within the four environments, lines in the AR09-192019 population consistently had significantly decreased average oil concentration, plant height, and lodging score, an earlier maturity date, and increased protein concentration compared to lines in the LD02-4485 population. These trends were similar to what were observed between the recurrent parents. Within environments, AR09-192019 lines had on average 17 to 29 g kg^{-1} more protein than LD02-4485 lines. Additionally, lines in the AR09-192019 population yielded less than those in the LD02-4485 population across and in each environment except for Pontiac 2016 where the two backgrounds were not significantly different from each other.

The QTL by genetic background interaction was non-significant ($P<0.05$) for all of the traits measured so further analyses were done across both genetic backgrounds. These analyses showed that lines containing the QTL allele from PI407788A had significantly greater average

protein concentration and less oil concentration (Table 3.2). Within environments, the mean increases in protein ranged from 8 to 14 g kg^{-1}, and across environments, the mean increase was 11 g kg^{-1}. Average yield and lodging score were not significantly different between the two genotypic classes; however, lines with the high protein allele had numerically lower average yields and lodging scores (Table 3.2; Table 3.3). Across environments, no significant difference was detected between lines with the high and low protein alleles for maturity date or plant height (Table 3.3). In Pontiac and Urbana 2016, lines with the PI407788A allele matured one day earlier ($P<0.05$) and were 3 cm and 2 cm shorter, respectively.

No statistically significant differences were observed for the effect of the PI407788A allele between the AR09-192019 and LD02-4485 backgrounds for all seed composition and agronomic traits when averaged across the four environments. In the AR09-192019 background, the high protein allele was associated with a 10 g kg^{-1} protein increase and a 109 kg ha^{-1} yield decrease. In the LD02-4485 background, the allele was associated with a 12 g kg^{-1} protein increase and a 57 kg ha^{-1} yield decrease.

Discussion

The high protein allele on chr 15 from PI407788A was tested across two MG II genetic backgrounds to determine its effect on seed composition and agronomic traits including yield. The QTL from PI407788A was mapped by Kim et al. (2015), but multiple seed composition QTL have been mapped to this location prior to and following that study (Diers et al., 1992; Sebolt et al., 2000; Fasoula et al., 2004; Pathan et al., 2013; Bandillo et al., 2015; Phansak et al., 2016; Qi et al., 2016.) Confirmation of this QTL in our study demonstrates its importance and further suggests that it is a stable candidate that can be used to develop cultivars with improved protein concentration. It also provides more evidence that diverse germplasm accessions such as PI407788A are useful sources to improve protein concentration in a breeding program.

The exponential decrease in genotyping costs and improvements in large data analysis capabilities have led to several recent protein and oil QTL mapping studies that used large and diverse population sizes and/or high density genetic markers (Hwang et al., 2014; Bandillo et al., 2015; Vaughn et al., 2015; Phansak et al., 2016; Qi et al., 2016). These studies have mapped additional seed composition QTL, but the chr 15 and chr 20 QTL are still consistently identified

as having the largest effects on protein and oil concentration. The effects of the chr 20 QTL on seed composition traits have generally been larger than those of the chr 15 QTL, which may account for the limited number of follow-up studies on the chr 15 QTL compared to the chr 20 QTL. Our results are consistent with that trend as the increases in protein concentration we observed for the chr 15 high protein allele were generally smaller than those previously reported for the chr 20 QTL (Sebolt et al., 2000; Chung et al., 2003; Nichols et al., 2006). Sebolt et al. (2000) attempted to characterize the effect of both QTL on agronomic traits but was unsuccessful in confirming the presence of the chr 15 QTL in their populations. To date, our study is the only successful attempt that we are aware of that tested the chr 15 QTL for its effect on both seed composition and agronomic traits.

Across genetic backgrounds, the high protein allele from PI407788A increased protein and decreased oil concentration, but it was not associated with significantly reduced yields (Table 3.2). This is in contrast to the evaluations of the chr 20 QTL and several other studies that observed a decrease in yield when protein was increased (Sebolt et al., 2000; Chung et al., 2003; Nichols et al., 2006). While the negative association between yield and protein concentration has been demonstrated in multiple studies, there is some evidence that the relationship can be separated when the right breeding scheme is applied (Brim and Burton, 1979; Holbrook et al., 1989; Wilcox and Cavins, 1995; Recker et al., 2014). Although we found no significant yield associations with the PI407788A allele, the yield of lines with the high protein allele were still numerically less than lines with the alternative alleles (Table 3.2).

Studies have shown that the chr 20 protein allele is associated with a yield reduction of 67 to 228 kg ha^{-1} for each 10 g kg^{-1} increase in protein (Sebolt et al., 2000; Nichols et al., 2006; Brzostowski , submitted). Although the chr 15 allele was not significantly associated with yield, and therefore, we need to be cautious in making conclusions due to the limited number of environments and genetic backgrounds evaluated, we estimate that this QTL is associated with a 69 kg ha^{-1} decrease in yield for each 10 g kg^{-1} increase in protein. This chr 15 yield association is on the lower end of the yield association observed for the chr 20 QTL, and more research is needed to verify whether the chr 15 could be use more successfully than the chr 20 QTL for increasing protein concentration without associated yield reductions.

Seed composition is heavily influenced by environmental conditions such as water availability and temperature, so it is possible that under different environmental conditions the

high protein allele could cause a significant yield decrease (Specht et al., 2001; Rotundo and Westgate, 2009). Also, genetics play a role in determining seed composition not only evidenced by the numerous seed composition QTL that have been mapped but also by the highly heritable nature of protein and oil concentration (Brummer et al., 1997; Chung et al., 2003; Eskandari et al., 2013). We did not observe a significant difference when we compared the genetic effects on each trait between the AR09-192019 and LD02-4485 backgrounds; however, this is only two backgrounds, and it is possible that significant difference in effects may be observed if tests were expanded to other genetic backgrounds. Therefore, it would be worthwhile to test the chr 15 QTL in other backgrounds and environments before it is deployed in a breeding program. Nonetheless, the stability of the genetic effects of the PI407788A allele across two backgrounds that were significantly different for all measured traits is promising (Table 3.1).

This study highlights the importance of evaluating the effect of high protein QTL on seed composition and agronomic traits. While it remains difficult to simultaneously increase protein and oil concentration and yield, the high protein allele from PI407788A could be a successful candidate for use in developing high protein soybean cultivars with economical seed yield. Additionally, the information presented in this study combined with new insights generated by rapidly evolving molecular technologies could help breeders form efficient and effective strategies to simultaneously improve seed composition and yield.

Tables

Table 3.1. Evaluation of seed composition and agronomic traits across lines in the AR09-192019 or LD02-4485 background. Protein and oil concentrations are reported on a 130 g kg^{-1} moisture basis.

	Seed yield (kg ha^{-1})				Protein (g kg^{-1})				Oil (g kg^{-1})				Maturity date†				Plant Height (cm)‡				Lodging (1-5)§			
Environment¶	AR09-192019#	LD02-4485††	Diff.‡‡	P value	AR09-192019	LD02-4485	Diff.	P value	AR09-192019	LD02-4485	Diff.	P value	AR09-192019	LD02-4485	Diff.	P value	AR09-192019	LD02-4485	Diff.	P value	AR09-192019	LD02-4485	Diff.	P value
Pontiac 2015	4913	5522	-609	<.0001	390	361	29	<.0001	212	220	-8	<.0001	907	913	-6	<.0001	82	90	-8	<.0001	1.3	1.6	-0.3	<.0001
Urbana 2015	4018	4556	-538	<.0001	372	343	29	<.0001	221	231	-10	<.0001	906	910	-4	<.0001	70	78	-8	<.0001	1.1	1.3	-0.2	<.0001
Pontiac 2016	4873	4886	-13	n.s.	386	362	24	<.0001	219	229	-10	<.0001	912	916	-4	<.0001	100	105	-5	<.0001	2.0	2.5	-0.5	<.0001
Urbana 2016	4612	4859	-247	<.0001	382	365	17	<.0001	222	229	-7	<.0001	908	912	-4	<.0001	91	99	-8	<.0001	1.4	1.9	-0.5	<.0001
Across	4603	4957	-354	<.0001	382	358	25	<.0001	219	228	-9	<.0001	908	912	-4	<.0001	86	93	-7	<.0001	1.4	1.8	-0.4	<.0001

n.s. = non-significant.

†The date when 95% of pods have reached mature color (R8; Fehr et al., 1971) with September 1 equivalent to 901.

‡Distance between the soil line and apex of the main stem.

§Lodging score was on a 1 to 5 scale with 1=all plants erect and 5=all plants prostrate.

¶Location and year combination.

Mean of all near-isogenic lines with AR09-192019 as their recurrent parent.

†† Mean of all near-isogenic lines with LD02-4485 as their recurrent parent.

‡‡ Difference between the means of lines in the AR09-192019 and LD02-4485 backgrounds.

Table 3.2. Average effects of the high protein allele from PI407788A on seed composition and yield across both the AR09-192019 and LD02-4485 backgrounds. Protein and oil concentration are reported on a 130 g kg^{-1} moisture basis.

	Seed yield (kg ha^{-1})				Protein (g kg^{-1})				Oil (g kg^{-1})			
Environment†	Donor Allele‡	Recurrent Allele§	Diff.¶	P value	Donor Allele	Recurrent Allele	Diff.	P value	Donor Allele	Recurrent Allele	Diff.	P value
Pontiac 2015	5226	5375	-151	n.s.	379	365	14	<.0001	214	221	-7	<.0001
Urbana 2015	4313	4407	-94	n.s.	360	347	13	<.0001	224	231	-7	<.0001
Pontiac 2016	4863	4899	-36	n.s.	376	366	10	<.0001	222	228	-6	<.0001
Urbana 2016	4749	4789	-40	n.s.	375	367	8	<.0001	224	229	-5	<.0001
Across	4790	4866	-76	n.s.	372	361	11	<.0001	221	227	-6	<.0001

n.s. = non-significant.

† Location and year combination.

‡ Mean of near-isogenic lines predicted to be homozygous for the PI407788A allele at the chr 15 locus based on the genetic marker Satt384.

§ Mean of near-isogenic lines predicted to be homozygous for the recurrent allele at the chr 15 locus based on the genetic marker Satt384.

¶ Difference between the means of homozygous classes.

Table 3.3. Average effects of the high protein allele from PI407788A on maturity date, plant height, and lodging across both the AR09-192019 and LD02-4485 backgrounds.

	Maturity date†				Plant Height (cm)‡				Lodging (1-5)§			
Environment¶	Donor Allele#	Recurrent Allele††	Diff.‡‡	P value	Donor Allele	Recurrent Allele	Diff.	P value	Donor Allele	Recurrent Allele	Diff.	P value
Pontiac 2015	911	912	-1	n.s.	87	88	-1	n.s.	1.4	1.5	-0.1	n.s.
Urbana 2015	908	909	-1	n.s.	74	75	-1	n.s.	1.2	1.2	0.0	n.s.
Pontiac 2016	913	914	-1	0.0263	102	105	-3	0.0186	2.3	2.4	-0.1	n.s.
Urbana 2016	910	911	-1	0.0129	95	97	-2	0.0495	1.7	1.8	-0.1	n.s.
Across	911	912	-1	n.s.	90	91	-2	n.s.	1.6	1.7	-0.1	n.s.

n.s. = non-significant.

†The date when 95% of pods have reached mature color (R8; Fehr et al., 1971) with September 1 equivalent to 901.

‡Distance between the soil line and apex of the main stem.

§ Lodging score was on a 1 to 5 scale with 1=all plants erect and 5=all plants prostrate.

¶Location and year combination.

Mean of lines predicted to be homozygous for the PI407788A allele at the chr 15 locus based on the genetic marker Satt384.

†† Mean of lines predicted to be homozygous for the recurrent allele at the chr 15 locus based on the genetic marker Satt384.

‡‡ Difference between the means of homozygous classes.

References

Abney, T.S. and W.D. Crochet. 2006. Uniform soybean tests northern states, 2011. USDA-ARS, West Lafayette, IN.

Abney, T.S. and T. Hughes. 2011. Uniform soybean tests northern states, 2011. USDA-ARS, West Lafayette, IN.

Anand, S.C., J.G. Shannon, J.A. Wrather, P.R. Arelli, D.A. Sleper, and L.D. Young. 2004. Registration of S97-1688 soybean germplasm line high in protein content and resistant to soybean cyst nematode. Crop Sci. 44: 698-700.

Bandillo, N., D. Jarquin, Q. Song, R. Nelson, P. Cregan, J. Specht, and A. Lorenz. 2015. A population structure and genome-wide association analysis on the USDA soybean germplasm collection. The Plant Genome. 8.

Bell-Johnson, B., G. Garvey, J. Johnson, D. Lightfoot, and K. Meksem. 1998. Biotechnology approaches to improving resistance to SCN and SDS: Methods for high throughput marker assisted selection. Soybean Genet. Newsl. 25:115-117.

Bolon, Y.T., B. Joseph, S.B. Cannon, M.A. Graham, B.W. Diers, A.D. Farmer, G.D. May, G.J. Muehlbauer, J.E. Specht, Z.J. Tu and N. Weeks. 2010. Complementary genetic and genomic approaches help characterize the linkage group I seed protein QTL in soybean. BMC Plant Bio. 10:41.

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

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

Brzostowski, L.F. 2017. Breeding for increased protein concentration, sudden death syndrome resistance, and soybean cyst nematode resistance in soybean. Ph.D. diss., Univ. of Illinois, Urbana.

Burton J, T. Carter, R. Wilson. 1999. Registration of 'Prolina' soybean. Crop Sci. 39:294-295.

Carter T.E., P.E. Rzewnicki, J.W. Burton, M.R. Villagarcia, D.T. Bowman , E. Taliercio, P. Kwanyuen. 2010. Registration of N6202 soybean germplasm with high protein, favorable yield potential, large seed, and diverse pedigree. J. Plant Reg. 4:73-79

Chen P, T. Ishibashi, D.G. Dombek, J.C. Rupe. 2011. Registration of R05-1415 and R05-1772 high-protein soybean germplasm lines. J. Plant. Reg. 5:410–413.

Cregan, P., and C. Quigley. 1997. Simple sequence repeat DNA marker analysis. In: G. Caetano-Anolles and Gresshoff, P.M. (eds), DNA markers: Protocols, applications. and overviews, John Wiley & Sons, New York. p. 173-185.

Cromwell, G.L. 2012. Soybean meal: An exceptional protein source. Soybean Meal InfoCenter, Ankeny, IA: available at <http://www.soymeal.org/ReviewPapers/SBMExceptionalProteinSource.pdf/>. Accessed January 30, 2017.

Chung, J., H.L. Babka, G.L Graef, P.E. Staswick, D.J. Lee, P.B. Cregan, R.C. Shoemaker, and J.E. Specht. 2003. The seed protein, oil, and yield QTL on soybean linkage group I. Crop Sci. 43:1053-1067.

Diers, B. W., P. Keim, W.R. Fehr, and R.C. Shoemaker. 1992. RFLP analysis of soybean seed protein and oil content. Theor. Appl. Genet. 83:608-612.

Eskandari, M., E.R. Cober, and I. Rajcan. 2013. Genetic control of soybean seed oil: II. QTL and genes that increase oil concentration without decreasing protein or with increased seed yield. Theor. Appl. Genet. 126:1677-1687.

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

Fehr, W., C. Caviness, D. Burmood, and J. Pennington. 1971. Stage of development descriptions for soybeans, *Glycine Max* (L.) Merrill. Crop Sci. 11:929-931.

Hartwig, E.E., T.M. Kuo, and M.M. Kenty. 1997. Seed protein and its relationship to soluble sugars in soybean. Crop Sci. 37:770-773.

Holbrook, C.C., J.W. Burton, and T.E. Carter. 1989. Evaluation of recurrent restricted index selection for increasing yield while holding seed protein constant in soybean. Crop Sci. 29:324-329.

Hwang, E.Y., Q. Song, G. Jia, J.E. Specht, D.L. Hyten, J. Costa, and P.B. Cregan. 2014. A genome-wide association study of seed protein and oil content in soybean. BMC Genomics. 15:1.

Hymowitz, T., F.I. Collins, J. Panczner, and W.M. Walker. 1972. Relationship between the content of oil, protein, and sugar in soybean seed. Agron. J. 64:613-616.

Keim, P., T. Olson, and R. Shoemaker. 1988. A rapid protocol for isolating soybean DNA. Soybean Genet. Newslet. 12:150-152.

Kim, M., S. Schultz, R. Nelson, and B.W. Diers. 2015. Identification and fine mapping of a soybean seed protein QTL from PI 407788A on chromosome 15. Crop Sci. 56:219-225.

Lee, S.H., M.A. Bailey, R.M. Mian, T.E. Carter, E.R. Shipe, D.A. Ashley, W.A. Parrot, R.S. Hussey, and H.R. Boerma. 1996. RFLP loci associated with soybean seed protein and oil content across populations and locations. Theor. Appl. Genet. 93:649-657.

Leffel, R.C., and W.K. Rhodes. 1993. Agronomic performance and economic value of high-seed-protein soybean. J. Prod. Agric. 6:365-368.

Mian, R., L. McHale, Z. Li, and A.E. Dorrance. 2017. Registration of 'Highpro1' soybean with high protein and high yield developed from a north × south cross. J. Plant. Reg. 11:51-54.

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

Panthee D.R., and V.R. Pantalone. 2006. Registration of soybean germplasm lines TN03–350 and TN04–5321 with improved protein concentration and quality registration by CSSA. Crop Sci. 46:2328-2329.

Pathan, S.M., T. Vuong, K. Clark, J. Lee, J. G. Shannon, C. A. Roberts, M.R. Ellersieck, J.W. Burton, P.B. Cregan, D.L. Hyten, H.T. Nguyen, and D.A. Sleper. 2013. Genetic mapping and confirmation of quantitative trait loci for seed protein and oil contents and seed weight in soybean. Crop Sci. 53:765-774.

Phansak, P., W. Soonsuwon, D.L. Hyten, Q. Song, P.B. Cregan, G.L. Graef, and J.E. Specht. 2016. Multi-population selective genotyping to identify soybean (*Glycine max* (L.) Merr.) seed protein and oil QTLs. G3: Genes| Genomes| Genetics. 6:1635-1648.

Qi, Z., J. Pan, X. Han, H. Qi, D. Xin, W. Li, X. Mao, Z. Wang, H. Jiang, C. Liu., and Z. Hu. 2016. Identification of major QTLs and epistatic interactions for seed protein concentration in soybean under multiple environments based on a high-density map. Mol. Breed. 36:1-16.

Recker, J.R., J.W. Burton, A. Cardinal, and L. Miranda. 2014. Genetic and phenotypic correlations of quantitative traits in two long-term randomly mated soybean populations. Crop Sci. 54:939-943.

Rotundo, J. L., and M.E. Westgate. 2009. Meta-analysis of environmental effects on soybean seed composition. Field Crops Res. 110:147-156.

SAS Institute. 2016. The SAS system for Microsoft Windows. Release 9.4. SAS Institute, Cary, NC.

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

SoyBase and the Soybean Breeder's Toolbox. Soybean breeder's toolbox genetic map information, available at <http://www.soybase.org>. Accessed January 19, 2017.

Specht, J.E., K. Chase, M. Macrander, G.L. Graef, J. Chung, J.P. Markwell, M. Germann, J.H. Orf, and K.G. Lark. 2001. Soybean Response to Water. Crop Sci. 41:493-509.

Vaughn, J.N., R.L. Nelson, Q. Song, P.B. Cregan, and Z. Li. 2014. The genetic architecture of seed composition in soybean is refined by genome-wide association scans across multiple populations. G3: Genes| Genomes| Genetics. 4:2283-2294.

Wang, D., J. Shi, S. Carlson, P. Cregan, R. Ward, and B. Diers. 2003. A low-cost, high-throughput polyacrylamide gel electrophoresis system for genotyping with microsatellite DNA markers. Crop Sci. 43:1828-1832.

Wilcox, J.R., and J.F. Cavins. 1995. Backcrossing high seed protein to a soybean cultivar. Crop Sci. 35:1036-1041.

Wilcox, J.R., and R.M. Shibles. 2001. Interrelationships among seed quality attributes in soybean. Crop Sci. 41:11-14.

Wilson, R.F. 2004. Seed composition. In: Boerma, H.R., and Specht, J.E. (eds.) Soybeans: improvement, production, and uses. 3rd ed. Agron. Monogr. 16. ASA, CSSA, and SSSA, Madison, WI.

Wilson, R.F. 2008. Soybean: Market driven research needs. p. 3-14. In Stacey, G. (ed.) Genetics and genomics of soybean. 2nd ed. Springer, New York, NY.

CHAPTER 4: FIELD EVALUATION OF THREE SOURCES OF GENETIC RESISTANCE TO SUDDEN DEATH SYNDROME OF SOYBEAN

Abstract

Sudden death syndrome (SDS) of soybean [*Glycine max* (L.) Merrill] is a disease that causes yield loss in soybean growing regions across the USA and worldwide. The most effective way to control the disease is through genetic resistance. While multiple quantitative trait loci (QTL) for SDS resistance have been mapped, studies to further evaluate these QTL are limited. The objective of our research was to test the effect of previously mapped SDS resistance QTL on foliar symptoms when incorporated into elite soybean backgrounds. Six backcross populations were developed to test the following QTL: cqSDS-001, with resistance originating from PI567374, a QTL on chromosome 10 from Ripley, and QTL on chromosomes 1 and 18 from PI507531. The NIL populations segregated for QTL alleles and were field tested in multiple inoculated environments and evaluated for leaf scorch. While foliar disease development was variable across environments and populations, each QTL was detected within at least one environment. This includes the detection of cqSDS-001 in three genetic backgrounds. The QTL allele from the resistant parent was associated with greater resistance than the susceptible allele for all QTL and backgrounds with the exception of the allele for the QTL on chromosome 18, where the opposite occurred. This study highlights the importance and difficulties of confirming QTL and the need for multi-year SDS field testing provides information for breeders to use to improve resistance to SDS.

Introduction

Sudden death syndrome (SDS) is an economically important disease of soybean [*Glycine max* (L.) Merr.] caused in North America by the soil-borne fungus, *Fusarium virguliforme* O'Donnell and T. Aoki (Aoki et al., 2003). It was first observed in Arkansas in 1972 and has since spread to almost all soybean growing regions in the US and as far north as Canada (Rupe et al., 2001). Additionally, the disease has been observed in Argentina and Brazil, the major soybean growing countries in South America (Ploper, 1993; Nakajima et al., 1993). The causal agents of SDS in South America are *F. virguliforme, F. tucumaniae, F. brasiliense, and F. crassistipitatum* (Aoki et al, 2003; 2005; 2012). Yield losses up to 80 percent have been reported in fields affected by SDS, but losses between 5 percent and 15 percent are more common (Roy et al., 1997). From 2007 to 2009, SDS was the fifth most important soybean disease in the USA with estimated yield losses ranging from approximately 0.56 to 0.94 million metric tons (Koenning and Wrather, 2010).

Foliar symptoms of SDS commonly develop when the plant is in its reproductive phase, and they begin as small scattered, interveinal light green or chlorotic spots resulting in a mottled appearance (Roy et al., 1997; Hartman et al., 2015). The spots then enlarge and can become necrotic or may run together to form larger areas of interveinal leaf cholorosis. In severe cases, defoliation occurs with petioles remaining attached to the stem. Root symptoms can also be associated with SDS and become more pronounced when foliar symptoms are severe. Roots of infected plants can exhibit crown necrosis and lateral root rot with gray or red brown discoloration radiating out from the pith. Blue sporulation may also be seen on the taproot and lower stem.

While management practices to protect against losses due to SDS are limited, one of the most effective ways to manage the disease is through genetic resistance. Field resistance to SDS has been classified as horizontal, rate reducing, and partial, it is controlled by many genes, and high heritabilities of over 0.80 for SDS resistance have been reported (Njiti et al., 1996; Njiti et al., 1997; Wen et al., 2014; Bao et al., 2015). Resistant genotypes have been identified and made available to producers; however, all genotypes will display some SDS symptoms if conditions are favorable (Hartman et al., 1997; Iqbal et al., 2001; Mueller et al., 2002; Mueller et al., 2003).

SDS resistance is complex, and it remains important for breeders to further their understanding of genetic resistance in order to develop better cultivars for producers.

Quantitative trait loci controlling SDS resistance have been mapped to over 11 chromosomes, and they are listed at: http://www.soybase.org/search/index.php?qtl=SDS (Soybase, 2017). Research has shown that QTL confer resistance to foliar or root injury or both (Njiti et al., 1998a; Triwotayakorn et al., 2005; Kazi et al., 2008; Abdelmajid et al., 2012), and only a few QTL have been mapped to the same regions across mapping populations (Njiti et al., 2002; Luckew et al., 2013; Wen et al., 2014; Bao et al., 2015; Zhang et al., 2015). In a population developed from a cross between 'Ripley' (Cooper et al., 1990) and 'Spencer'(Wilcox et al., 1990), de Farias Neto et al. (2007) mapped resistance QTL on chromosome (chr) 19 (formally linkage group (LG) L) and chr 17 (LG D2). The QTL on chr 17 mapped to the same genetic region in a population derived from crossing PI567374 by 'Omaha' (Nickell et al., 1998) and was confirmed and designated cqSDS-001. This is the only QTL that has been confirmed based on the rules set by the Soybean Genetics Committee (http://www.soybase.org/resources/QTL.php). Most SDS resistance QTL have been mapped in biparental populations; however, advances in genotyping and statistical analysis have led to recent genome wide association studies that have identified novel and previously identified SDS resistance QTL (Wen et al., 2014; Bao et al., 2015; Zhang et al., 2015).

After QTL have been mapped, follow-up research to better understand these QTL is needed. It is important to confirm and incorporate QTL into multiple genetic backgrounds to determine whether they will maintain their effect and be useful in breeding programs (Fasoula et al., 2004). A QTL that does not maintain its desired effect across multiple backgrounds could be dependent on a mechanism specific to a genotype and therefore will have limited use to a breeder. The objective of this study is to evaluate previously mapped SDS resistance QTL from PI567374, Ripley, and PI507531 in backcross soybean populations.

Materials and Methods

Population Development

Six segregating populations were developed via backcrossing to incorporate four SDS resistance QTL separately into different genetic backgrounds (Table 4.1).

The QTL allele on chromosome 17 (cqSDS-001) from PI567374 (de Farias Neto et al., 2007) was backcrossed four generations into three genetic backgrounds to form three populations of BC_4F_2-derived lines. The backgrounds were adapted to central Illinois and had SCN resistance from PI88788 or PI437654 (Cary and Diers, 2004). They included LD02-4485 (Abney and Crochet, 2006), a maturity group (MG) II cultivar; LD01-5907 (Abney and Crochet, 2006), a late MG III experimental line; and LD00-3309 (Diers et al., 2006), a MG IV cultivar.

The QTL allele on chromosome 10 from Ripley (Pruski, personal communication) was backcrossed four generations into LD02-4485 to create a population consisting of BC_4F_2-derived lines. The resistance QTL from Ripley will herein be referred to as CHR10.

Two additional populations of BC_3F_2-derived lines were developed by individually backcrossing resistance alleles for QTL on chromosomes 1 and 18 from PI507531 (Pruski, personal communication) into Spencer, a MG IV cultivar highly susceptible to SDS. The resistance QTL from PI507531 on chromosomes 1 and 18 will herein be referred to as CHR1 and CHR18, respectively.

Polymorphic markers within and flanking the genetic regions of interest were used to perform marker-assisted selection (MAS) and introgress each QTL region into one or more recurrent parents. The approximate size of the introgressed intervals are as follows: cqSDS-001, 12.5 Mb; CHR 10, 7.7 Mb; CHR1, 45.9 Mb; CHR18, 33.8 Mb. Once backcrossing was completed, heterozygous BC_nF_2 plants were self-pollinated to form populations of BC_nF_2-derived lines, and lines homozygous for the QTL allele from the resistant or susceptible parent were selected based on flanking markers. Lines that had recombinations between markers were discarded.

Field Evaluation

Each population was evaluated for three growing seasons at inoculated locations between 2012 and 2015 (Table 4.1). Early maturing populations were planted at locations in Ames, IA; Decatur, MI; and Urbana, IL. Late maturing populations were planted at locations in Urbana; Valmeyer; and Shawneetown, IL. In 2012, planting dates were as follows: Ames, April 18; Decatur, May 6; Shawneetown, n/a; Urbana, April 19; Valmeyer, n/a. In 2013, planting dates were as follows: Decatur, April 30; Shawneetown, n/a ; Urbana, May 14; Valmeyer, n/a. In 2014, planting dates were as follows: Decatur, May 5; Shawneetown, n/a; Urbana, April 24; Valmeyer, n/a. In 2015, planting dates were as follows: Decatur, May 2; Shawneetown, n/a; Urbana, April 29; Valmeyer, n/a. Plant inoculum in the form of *F. virguliforme* infested sorghum kernels was incorporated into the seedbed at most Iowa and Illinois locations (de Farias Neto et al., 2006). The inoculum rate was as follows: 14 cm^3/1 m of row at Ames and 34 cm^3/1 m of row at the inoculated Illinois locations. The Decatur and Valmeyer locations were naturally infested with *F. virguliforme*. Lines from each population plus recurrent parents and checks were blocked separately, and plots were arranged in a randomized completed block design. Plot length, row spacing, and seeding rate were variable. At Urbana, two-row plots were 3.05 m long with 0.76 m row spacing and a seeding rate of ~30 seeds per m. Plot length, row spacing, and seeding rate were similar at Decatur except plots were a single row. At Ames, two-row plots were 2.40 m long with a 0.80 m row spacing and a seeding rate of ~27 seeds per m (Table 4.1). Entries were replicated twice at Ames; Decatur; and Urbana, and three times at Valmeyer and Shawneetown. During reproductive growth stages, locations were irrigated as needed using drip tape or a center pivot irrigation system to promote SDS symptom development.

Foliar disease incidence and severity readings were taken on a per plot basis at approximately the R6 growth stage (Fehr et al., 1971). Disease incidence (DI) was measured as a percentage of plants in a plot displaying foliar SDS symptoms. Foliar disease severity (DS) was rated on a scale from 1 to 9 according to Njiti et al (1996): 1 = 0-10% of leaf surface chlorotic or 1-5% necrotic, 2 = 10-20% of leaf surface chlorotic or 6-10% necrotic, 3 = 20 – 40% of leaf surface chlorotic or 10-20% necrotic, 4 = 40-60% of leaf surface chlorotic or 20-40% necrotic, 5 = > 60% of leaf surface chlorotic or > 40% necrotic, 6 = up to 33% premature defoliation, 7 = up to 66% premature defoliation, 8 = > 66% premature defoliation, 9 = premature death of the plant.

Disease index (DX) was calculated with a possible range of 0 to 100 using the following equation: DX=(DS x DI)/9 (Njiti et al., 1996).

Genetic Analysis

Genomic DNA was isolated from expanding young trifoliate leaves using either a modified CTAB (hexadecyltrimethylammonium-bromide) method (Keim et al., 1988) or a modified quick DNA extraction method (Bell-Johnson et al., 1998). Polymerase chain reactions (PCR) were done according to Cregan and Quigley (1997) to test the NILs with simple sequence repeat (SSR) markers (Cregan et al., 1999; Song et al., 2004; Song et al., 2010). Electrophoresis was used to separate PCR products on 6% (w/v) nondenaturing polyacrylamide gels (Wang et al., 2003).

Statistical Analysis

The PROC MIXED function in SAS v9.4 (SAS Institute Inc., Cary, NC) was used to analyze leaf scorch symptoms in each population. Environments were defined as year by location combinations. Marker genotype, line, and marker genotype nested within line were fixed, and environment and replication were random. Interactions that included a random effect were also random. Data were transformed using the square root function if the normality assumption of the residuals was not met. Additionally, in the case of heterogeneous error variances, the mixed model was fitted using the REPEATED statement with GROUP=option (Littell et al., 2006). Reported p values are based on transformed data whereas all other data have been back transformed to original units.

Results

Field Test Overview

The soybean populations were developed to be representative of the maturity groups and cultivars planted in Illinois. The six populations were evaluated at multiple locations over three

growing seasons (Table 4.1). Foliar symptom development varied according to the environment and population. In several environments, limited disease expression or the presence of other pathogens made it impossible to record leaf scorch ratings.

cqSDS-001

QTL cqSDS-001 on chromosome 17 was tested in the LD02-4485, LD01-5907, and LD00-3309 backgrounds, and the populations consisted of 36, 44, and 36 lines, respectively (Table 4.1). For the LD02-4485 population, foliar disease data were not collected in the 2014 Decatur, MI environment due to inadequate disease development (Table 4.2). In 2012, lines with the resistance QTL allele had a significantly ($P<0.05$) reduced average DS and DX at Decatur and a reduced average DX at Urbana. In Urbana 2013, the average DX and DI of lines containing the resistance allele at cqSDS-001 was significantly less than those containing the susceptible allele. Although these significant differences were observed within multiple environments, they were small in magnitude, which is likely the result of the foliar symptoms being weak in the LD02-4485 background. The effect of the QTL was not observed across environments.

For the LD01-5907 population, SDS foliar symptoms were recorded at four environments (Table 4.3). No data were collected from this population in 2015 or at the Shawneetown location in 2013 and 2014 as there was insufficient symptom development. Lines homozygous for resistance at the cqSDS-001 locus had a significantly lower DX, DI, and DS at Valmeyer in 2014. Although the difference in DS was only a fraction of a rating point, the difference in DI was over 13%. The resistance QTL allele's effect was not detected across environments.

Leaf scorch ratings were collected at six environments for the LD00-3309 population (Table 4.4). Reliable symptoms did not develop at the Shawneetown location during the three field seasons of evaluation. Even though symptom expression was light, significant differences among lines were observed at Urbana in 2012 and 2013. In Urbana 2012, lines homozygous for the PI567374 allele had a reduced average DI and DX compared to lines homozygous for the LD00-3309 allele. In Urbana 2013, the average DS, DI, and DX of lines with the resistance allele was reduced compared to lines with the LD00-3309 allele. Despite being statistically significant,

all differences observed in the LD00-3309 population were small in magnitude. In the LD00-3309 background, cqSDS-001 was not significant across environments.

Chr 10 QTL from Ripley (CHR10)

CHR10 was evaluated in the LD02-4485 background in a population of 48 NILs (Table 4.1). The population was tested in six environments, and there was inadequate disease development for data collection at the 2015 Urbana location (Table 4.5). With the exception of 2013 Decatur, MI where symptoms were moderate, foliar disease development was low. The only environment that a significant difference between lines containing the resistance QTL allele and the susceptible allele was Urbana 2014. Despite low level disease expression, lines with the resistance allele at CHR10 had an average DS of 0.34 while lines with the susceptible allele had an average DS of 0.70. No significant QTL effect was observed across environments.

Chr 18 QTL from PI507531 (CHR18)

Using Spencer as the recurrent parent, a population of 35 lines was developed to determine the effect of CHR18 (Table 4.1). This population was planted at nine environments; however, ratable foliar symptoms were only observed at six environments (Table 4.6). Spencer is susceptible to many foliar diseases in addition to SDS, which made visual data collection impossible at the three nonrated environments. While foliar symptoms were moderate to severe in most environments, a significant effect for CHR18 was only detected in 2015 at Shawneetown. Unfortunately, the QTL effect was opposite of what was expected with the average DX significantly greater for lines homozygous for the resistance allele compared to homozygous susceptible lines; however, the difference was relatively small. This unexpected effect was supported by a similar, but not significant, trend of in other environments. Across environments, no significant differences due to the CHR18 QTL were detected.

Chr 1 QTL from PI507531 (CHR1)

A population consisting of 46 lines was developed to test the effect of CHR1 on SDS foliar symptoms in the Spencer background (Table 4.1). Adequate leaf scorch symptoms were observed in eight out of the nine environments where the population was planted (Table 4.7). Overall, disease development was robust in the eight environments where data were collected, and symptoms were moderate to severe.

CHR1 was detected within and across environments; however, the average difference between resistant and susceptible lines tended to be small especially for DS. Within the 2013 and 2014 Urbana, IL environments, lines with the CHR1 resistance alleles had a significantly lower average DS, DI, and DX than lines with the susceptible allele. In addition, lines with the resistance QTL allele had a significantly reduced average DS and DX within 2013 Valmeyer and a significantly reduced DX within 2014 Valmeyer. Across 2013 environments, the CHR1 resistance allele was associated with decreased DS while across 2014 environments the resistance allele was associated with decreased DX. Across all environments, there was a significant reduction in DS for lines containing the resistance allele compared to lines containing the susceptible allele.

Discussion

The objective of this study was to identify QTL breeders can use to successfully improve SDS resistance in their current programs. In this study, QTL effects on resistance were detected in at least one environment. With the exception of lines in the CHR18 population, lines with a resistance allele had a reduced average disease severity, disease incidence, and/or disease severity compared to lines with the susceptible allele. QTL effects for cqSDS-001 on resistance were detected in the LD02-4485, LD01-5907, and LD00-3309 backgrounds and within multiple environments. The CHR1 resistance allele from PI507531 significantly reduced SDS foliar symptoms within three environments and averaged across all environments compared to the susceptible allele. While differences between lines containing the resistance QTL alleles and lines containing the susceptible QTL alleles were observed, they were often minor. Significant differences between QTL classes for DS were a fraction of a point on the rating scale. It is

possible the QTL would have a greater effect on resistance in different genetic backgrounds or different environments; however, it is also possible that these QTL are not good candidates to increase SDS resistance in cultivars at a level that provides an economic benefit to producers. Additionally, cqSDS-001 and CHR10 were detected in populations and environments where symptom development was minor. Therefore, it is also possible that this QTL only improves resistance when disease pressure is low. Although we did not test for yield so we cannot estimate the impact of these QTL on yield, it has been shown previously that yield loss can still occur when symptoms are mild to moderate (Njiti et al., 1998b).

It can be challenging to achieve robust and reliable SDS foliar symptoms in the field. As in other SDS field studies, considerable effort was taken to optimize field environments to make them conducive to SDS development (de Farias Neto et al., 2006; de Farias Neto et al., 2007; Kandel et al., 2015; Weems et al., 2015). The plots were planted early, irrigated as needed during reproductive growth stages, and the soil was inoculated with *Fusarium virguliforme*; however, the foliar symptoms in some tests were still too light to rate or nonexistent. Environmental factors that are difficult to control in a field setting such as temperature, rainfall, and soil texture all play a major role in SDS disease development (Scherm and Yang, 1996; Roy et al., 1997; Scherm et al., 1998; Hartman et al., 2015).

The genetic backgrounds used in this study were selected because they are characteristic of high yielding, MG II-IV cultivars grown currently or previously across Illinois and other soybean growing states in the midwestern USA. LD02-4485, LD01-5907, and LD00-3309 are also moderately resistant to SDS, and this could have been another barrier to obtaining good foliar symptoms and an explanation for the minor significant differences observed between QTL classes within the populations. These three backgrounds have SCN resistance alleles at *rhg1*, and the genetic interval containing this allele has been shown to be associated with improved SDS resistance (Prabhu et al., 1999; Njiti et al., 2002; Triwitayakorn et al., 2005; Srour et al, 2012). For the most part, we were able to consistently achieve symptoms and collect leaf scorch data in the susceptible background, Spencer. In future SDS confirmation studies, it may be more effective to test QTL in high-yielding genetic backgrounds that have been previously shown to be moderately susceptible or susceptible to the disease.

In some environments, SDS symptoms were most likely present, but we could not collect reliable data due to the presence of other pathogens. Several other fungal diseases of soybean

including brown stem rot (causal agent-*Phialophora gregata*), red crown rot (*Cylindrocladium parasiticum*), charcoal rot (*Macrophomina phaseolina*), and stem canker (*Diaporthe phaseolorum* var. *meridionalis*, *Diaporthe phaseolorum* var. *caulivora*) cause foliar leaf scorch symptoms similar to SDS (Hartman et al., 2015). Destructive sampling is often the only way to tell the diseases apart, which is not necessarily feasible in field experiments with many locations, replications, and entries. Additionally, it can be common for plants to be infected with several other pathogens that produce symptoms that make the visual diagnosis of SDS difficult or impossible. This is especially the case when SDS symptoms appear late in the growing season around the R7 growth stage or at a lower position a plant's canopy. The Shawneetown location often had other foliar symptoms that confounded our SDS ratings. The complex nature of screening for resistance to SDS demonstrates the importance of extensive multi-environment field experiments to effectively breed for enhanced SDS resistance.

There have been rapid advances in genetics and statistical modeling, and there is hope that next generation breeding methods will lead to cultivars with durable resistance to SDS. Association mapping studies have mapped novel and previously mapped SDS resistance QTL across several diverse association panels (Wen et al., 2014; Bao et al., 2015; Zhang et al., 2015). They have also offered insight to help us better understand the complexity of SDS resistance; however, many questions still remain. Genomic selection (GS), which uses predictive models created with phenotypic and genotypic data to generate genomic estimated breeding values (GEBVs), is being tested for use in breeding for SDS resistance. While the reported prediction accuracies of GS models for SDS related traits have been low, GS still holds promise to increase SDS resistance (Bao et al., 2015). Research evaluating the implementation of GS to improve resistance to SDS is limited and still in its infancy. Although recent studies have suggested the potential of association mapping and genomic selection to increase durable resistance to SDS, more exploration is needed before these methods can be successfully applied in a breeding program. Additionally, until a reliable system is established, accurate phenotypic assessment will continue to impede the efficacy of breeding for improved resistance to SDS.

The QTL evaluated in this study could be used to pyramid resistance QTL into elite cultivars for more durable resistance. While the genetic mechanisms of the QTL we evaluated are unknown, stacking the two distinct SDS resistance mechanisms, resistance to root rot and leaf scorch, could also be a strategy to increase resistance. Although stacking multiple QTL into

a single cultivar is a time consuming endeavor, MAS improves efficiency and reduces the time needed to do so. In the future, predictive modeling in combination with high throughput genetic data could conserve the resources and time needed to incorporate multiple resistance QTL into elite cultivars. It would also help capture the resistance of multiple small effect QTL.

de Farias Neto et al. (2007) mapped and confirmed cqSDS0-001 using resistance originating from Ripley and also mapped a QTL to the same interval from PI567374. In our study, we have now detected the effect of resistance allele from PI567374 in three backgrounds and new environments. This shows that this QTL can increase resistance across backgrounds and environments and therefore could be useful for breeders as they breed for SDS resistance.

While the CHR10 allele from Ripley and the CHR1 and CHR18 alleles from PI507531 had been previously mapped, there has not been further evaluation of these QTL outside of the original study. We have observed the impact of the CHR1 and CHR10 QTL in new environments and backgrounds showing that positive effects of these resistance alleles. The CHR1 QTL showed a positive effect on resistance in four environments and across environments indicating that it could be promising for use in breeding programs. The increase in resistance was small, so more information would be needed to validate whether this QTL could improve resistance at the increased level necessary for resistant cultivars.

While we observed a significant effect for the CHR18 QTL interval, the QTL effect was opposite of expectations. There are a number of explanations for the unexpected result. This includes that (1) the QTL effect detected in the current study was not for the same QTL detected previously, but instead a closely linked QTL, (2) there are multiple alleles for the QTL and the donor parent had an allele for greater resistance than the PI parent, (3) the original QTL and the marker used to track it during backcrossing were not tightly linked and the resistance allele was lost during backcrossing, (4) the parents of the backcross population were not segregating for QTL alleles previously detected and the significance was a Type 1 error, and (5) there is not a QTL in this region and the significance in the mapping study was a Type 1 error. Overall, it is difficult to identify large effect QTL to improve traits such as SDS resistance because QTL detection and confirmation can be inconsistent, and these factors make it important to critically evaluate QTL prior to wide use in a breeding program (Beavis et al., 1994; Kao and Zang, 1997; Fasoula et al., 2004).

There are numerous challenges a breeder must overcome to improve resistance to SDS. These include difficulty in accurately evaluating symptoms and the complex genetic nature of resistance. Nonetheless, genetic resistance remains critical in limiting the yield loss caused by sudden death syndrome. Our study indicates there are several sources including PI567374 and PI507531 a breeder could use to enhance SDS resistance in their program. It is important to note that the significant increases in resistance associated with these two sources were small in magnitude and were only noted in a limited number of environments, and therefore, they could be of limited use to improve resistance to SDS. Nonetheless, the detection of resistance QTL effects from PI567374 within multiple genetic backgrounds and from PI507531 across environments is promising. This study also highlights both the necessity and complicated nature of multiyear field testing to evaluate symptoms, and the importance of evaluating QTL across multiple genetic backgrounds prior to incorporating them into a breeding program. It is expected that advances in high throughput genotyping and phenotyping will offer better insight into the complexity of sudden death syndrome and will ultimately result in improved cultivars for producers. In the meantime, implementing MAS using QTL such as cqSDS-001 and CHR1 and the continued evaluation of mapped QTL could be effective methods to mitigate losses due to this devastating disease.

Tables

Table 4.1. Description of the populations evaluated for sudden death syndrome foliar symptoms and the environments where they were evaluated.

Parentage of BC Population†	Generation‡	QTL§	MG¶	No. Entries#		Years††	Location‡‡				
				Res	Sus		Ames, IA	Decatur, MI	Shawneetown, IL	Urbana, IL	Valmeyer, IL
LD02-4485 x PI567374	$BC_4F_{2:n}$	cqSDS-001	2.6	19	17	2012, 2013, 2014	x	x		x	
LD02-4485 x Ripley	$BC_4F_{2:n}$	chr 10	2.6	19	29	2013, 2014, 2015		x		x	
LD01-5907 x PI567374	$BC_4F_{2:n}$	cqSDS-001	3.8	22	22	2013, 2014, 2015			x	x	x
Spencer x PI507531	$BC_3F_{2:n}$	chr 1	4.0	28	18	2013, 2014, 2015			x	x	x
Spencer x PI507531	$BC_3F_{2:n}$	chr 18	4.0	16	19	2013, 2014, 2015			x	x	x
LD00-3309 x PI567374	$BC_4F_{2:n}$	cqSDS-001	4.0	18	18	2012, 2013, 2014			x	x	x

†The recurrent parent (listed first) crossed to the donor parent (listed second).

‡Backcross generation of the population when it was evaluated.

§ The QTL name or chromosome (chr) location of the QTL segregating in the population.

¶ Relative maturity of the recurrent parent.

Number of lines in the test. “Res” indicates lines are predicted to be homozygous resistant based on markers linked to the QTL. “Sus” indicates lines are predicted to be homozygous susceptible based on markers linked to the QTL.

†† Years when population was evaluated.

‡‡ Location where population was planted. An “x” indicates the population was planted at a location.

Table 4.2. Evaluation of sudden death syndrome disease severity, disease incidence, and disease index in a population of lines segregating for the resistance QTL, cqSDS-001, in the LD02-4485 background.

		Disease Severity†				Disease Incidence‡				Disease Index§			
Year	Location¶	Res#	Sus††	Diff‡‡	P value§§	Res	Sus	Diff	P value	Res	Sus	Diff	P value
2012	Ames, IA	1.37	1.35	--	0.75	5.4	3.9	--	0.51	1.80	1.30	--	0.61
2012	Decatur, MI	0.45	0.84	-0.39	0.01	6.0	8.3	--	0.11	0.84	1.56	-0.72	0.04
2012	Urbana, IL	0.54	1.03	--	0.07	2.3	3.9	--	0.10	0.39	0.81	-0.42	0.04
2013	Decatur, MI	1.96	1.94	--	0.64	34.7	29.6	--	0.23	8.2	8.97	--	0.62
2013	Urbana, IL	0.11	0.39	-0.28	0.04	0.6	2.0	-1.4	0.04	0.12	0.37	--	0.07
2014	Urbana, IL	0.22	0.46	--	0.25	0.5	1.1	--	0.22	0.11	0.29	--	0.23

†Disease severity (DS) was rated on a 1-9 scale according to Njiti et al., 1996.

‡Disease incidence (DI) is the percent of the two-row plot displaying foliar symptoms.

§Disease index (DX) was calculated using the following equation: DX=((DS*DI)/9) (Njiti et al., 1996).

¶Location where disease ratings were recorded.

#Means of lines predicted to be homozygous resistant (Res) at cqSDS-001 based on the genetic markers Sat_222, Satt311, Satt301, Satt186.

†† Means of lines predicted to be homozygous susceptible (Sus) at cqSDS-001 based on the genetic markers Sat_222, Satt311, Satt301, Satt186.

‡‡Difference between the means of homozygous classes.

§§Probability value used to compare the mean of homozygous resistant and homozygous susceptible lines.

Table 4.3. Evaluation of sudden death syndrome disease severity, disease incidence, and disease index across lines segregating for the resistance QTL, cqSDS-001, in the LD01-5907 background.

		Disease Severity†				Disease Incidence‡				Disease Index§			
Year	Location¶	Res#	Sus††	Diff‡‡	P value§§	Res	Sus	Diff	P value	Res	Sus	Diff	P value
2013	Urbana, IL	0.81	1.10	--	0.22	3.3	4.1	--	0.33	0.74	0.91	--	0.28
2013	Valmeyer, IL	1.89	1.89	--	0.66	84.9	82.7	--	0.72	18.50	18.40	--	0.71
2014	Urbana, IL	0.74	0.69	--	0.82	1.8	2.2	--	0.90	0.32	0.49	--	0.60
2014	Valmeyer, IL	1.36	1.61	-0.25	0.04	68.9	82.2	-13.3	0.01	11.10	15.30	-4.20	0.003

†Disease severity (DS) was rated on a 1-9 scale according to Njiti et al., 1996.

‡Disease incidence (DI) is the percent of the plot displaying foliar symptoms.

§Disease index (DX) was calculated using the following equation: DX=((DS*DI)/9) (Njiti et al., 1996)

¶Location where disease ratings were recorded.

#Means of lines predicted to be homozygous resistant (Res) at cqSDS-001 based on the genetic markers Sat_222, Satt311, Satt488, Satt301, GMHSP179, Satt186, Satt031.

†† Means of lines predicted to be homozygous susceptible (Sus) at cqSDS-001 based on the genetic markers Sat_222, Satt311, Satt488, Satt301, GMHSP179, Satt186, Satt031.

‡‡Difference between the means of homozygous classes.

§§Probability value used to compare the mean of homozygous resistant and homozygous susceptible lines.

Table 4.4. Evaluation of sudden death syndrome disease severity, disease incidence, and disease index across lines segregating for the resistance QTL, cqSDS-001, in the LD00-3309 background.

		Disease Severity†				Disease Incidence‡				Disease Index§			
Year	Location¶	Res#	Sus††	Diff‡‡	P value§§	Res	Sus	Diff	P value	Res	Sus	Diff	P value
2012	Urbana, IL	0.75	1.10	--	0.21	1.1	2.3	-1.2	0.046	0.22	0.54	-0.32	0.04
2012	Valmeyer, IL	1.10	1.50	--	0.15	12.6	13.8	--	0.63	3.00	4.90	--	0.3
2013	Urbana, IL	0.60	1.20	-0.60	0.01	1.1	2.3	-1.2	0.01	0.25	0.49	-0.24	0.03
2013	Valmeyer, IL	2.10	2.30	--	0.30	97.7	97.6	--	0.94	23.40	24.40	--	0.38
2014	Urbana, IL	0.54	0.33	--	0.31	2.0	1.1	--	0.14	0.55	0.14	--	0.14
2014	Valmeyer, IL	1.22	1.24	--	0.87	69.4	66.9	--	0.61	10.00	12.50	--	0.84

†Disease severity (DS) was rated on a 1-9 scale according to Njiti et al., 1996.

‡Disease incidence (DI) is the percent of the plot displaying foliar symptoms.

§Disease index (DX) was calculated using the following equation: DX=((DS*DI)/9) (Njiti et al., 1996)

¶Location where disease ratings were recorded.

#Means of lines predicted to be homozygous resistant (Res) at cqSDS-001 based on the genetic markers Sat_222, Satt488, and Satt186.

†† Means of lines predicted to be homozygous susceptible (Sus) at cqSDS-001 based on the genetic markers Sat_222, Satt488, and Satt186.

‡‡Difference between the means of homozygous classes.

§§Probability value used to compare the mean of homozygous resistant and homozygous susceptible lines.

Table 4.5. Evaluation of sudden death syndrome disease severity, disease incidence, and disease index across lines segregating for the resistance QTL from Ripley on chromosome 10 in the LD02-4485 background.

		Disease Severity†				Disease Incidence‡				Disease Index§			
Year	Location¶	Res#	Sus††	Diff‡‡	P value§§	Res	Sus	Diff	P value	Res	Sus	Diff	P value
2013	Decatur, MI	2.90	2.80	--	0.69	47.8	49.7	--	0.75	16.50	17.00	--	0.79
2013	Urbana, IL	0.46	0.33	--	0.27	2.1	1.7	--	0.36	0.33	0.21	--	0.23
2014	Decatur, MI	0.10	0.29	--	0.11	1.6	3.1	--	0.32	0.18	0.91	--	0.19
2014	Urbana, IL	0.34	0.70	-0.36	0.04	1.0	1.7	--	0.07	0.23	0.43	--	0.14
2015	Decatur, MI	0.86	0.66	--	0.19	18.7	13.7	--	0.18	2.66	1.92	--	0.17

†Disease severity (DS) was rated on a 1-9 scale according to Njiti et al., 1996.

‡Disease incidence (DI) is the percent of the plot displaying foliar symptoms.

§Disease index (DX) was calculated using the following equation: DX=((DS*DI)/9) (Njiti et al., 1996)

¶Location where disease ratings were recorded.

#Means of lines predicted to be homozygous resistant (Res) at the QTL locus on chromosome 10 based on the genetic markers Satt653 and Satt345.

†† Means of lines predicted to be homozygous susceptible (Sus) at the QTL locus on chromosome 10 based on the genetic markers Satt653 and Satt345.

‡‡Difference between the means of homozygous classes.

§§Probability value used to compare the mean of homozygous resistant and homozygous susceptible lines.

Table 4.6. Evaluation of sudden death syndrome disease severity, disease incidence, and disease index across lines segregating for the resistance QTL from PI507531 on chromosome 18 in the 'Spencer' background.

		Disease Severity†				Disease Incidence‡				Disease Index§			
Year	Location¶	Res#	Sus††	Diff‡‡	P value§§	Res	Sus	Diff	P value	Res	Sus	Diff	P value
2013	Urbana, IL	3.80	3.40	--	0.10	53.4	45.8	--	0.11	23.70	18.70	--	0.10
2013	Valmeyer, IL	4.60	4.40	--	0.39	100.0	100.0	--	1.00	50.70	49.10	--	0.40
2014	Urbana, IL	2.40	2.20	--	0.60	32.6	28.7	--	0.60	12.30	9.70	--	0.39
2014	Valmeyer, IL	3.00	3.17	--	0.37	98.0	98.3	--	0.89	32.60	35.00	--	0.38
2015	Urbana, IL	1.73	1.57	--	0.62	46.0	43.0	--	0.86	13.84	11.54	--	0.49
2015	Shawneetown, IL	6.07	5.90	--	0.25	62.5	57.4	--	0.21	42.00	37.03	5.0	0.04

†Disease severity (DS) was rated on a 1-9 scale according to Njiti et al., 1996.

‡Disease incidence (DI) is the percent of the plot displaying foliar symptoms.

§Disease index (DX) was calculated using the following equation: DX=((DS*DI)/9) (Njiti et al., 1996)

¶Location where disease ratings were recorded.

#Means of lines predicted to be homozygous resistant (Res) at the QTL locus on chromosome 18 based on the genetic markers Satt115, Satt566, and Satt504.

†† Means of lines predicted to be homozygous susceptible (Sus) at the QTL locus on chromosome 18 based on the genetic markers Satt115, Satt566, Satt504 .

‡‡Difference between the means of homozygous classes.

§§Probability value used to compare the mean of homozygous resistant and homozygous susceptible lines.

Table 4.7. Evaluation of sudden death syndrome disease severity, disease incidence, and disease index across lines segregating for the resistance QTL from PI507531 on chromosome 1 in the 'Spencer' background.

		Disease Severity†				Disease Incidence‡				Disease Index§			
Year	Location¶	Res#	Sus††	Diff‡‡	P value§§	Res	Sus	Diff	P value	Res	Sus	Diff	P value
2013	Urbana, IL	3.40	4.10	-0.70	0.015	44.6	54.2	-9.6	0.04	18.90	26.30	-7.40	0.009
2013	Valmeyer, IL	4.00	4.44	-0.44	0.004	100	100	--	1.00	44.20	49.30	-5.10	0.004
2013	Shawneetown, IL	0.82	0.86	--	0.45	16.8	11.0	--	0.35	2.70	2.03	--	0.50
2014	Urbana, IL	2.10	2.70	-0.60	0.03	31.4	40.0	-8.6	0.048	10.20	14.40	-4.20	0.05
2014	Valmeyer, IL	2.60	2.90	--	0.06	92.7	96.0	--	0.28	27.10	31.40	-4.30	0.03
2015	Urbana, IL	2.96	2.85	--	0.58	48.5	47.8	--	0.995	21.90	20.35	--	0.78
2015	Valmeyer, IL	2.33	2.39	--	0.50	86.9	89.8	--	0.51	23.61	24.76	--	0.52
2015	Shawneetown, IL	6.18	6.18	--	0.91	58.3	58.1	--	0.95	12.10	11.59	--	0.74
Across 2013		2.70	3.10	-0.40	0.004	53.80	54.90	--	0.77	22.00	25.80	--	0.31
Across 2014		2.37	2.79	--	0.30	62.50	67.80	--	0.43	18.80	22.90	-4.10	0.005
Across 2015		2.26	2.24	--	0.96	65.12	65.98	--	0.81	18.92	18.77	--	0.99
Across 2013, 2014, and 2015		2.47	2.71	-0.24	0.02	60.19	62.04	--	0.26	20.14	22.53	--	0.07

†Disease severity (DS) was rated on a 1-9 scale according to Njiti et al., 1996.

‡Disease incidence (DI) is the percent of the plot displaying foliar symptoms.

§Disease index (DX) was calculated using the following equation: DX=((DS*DI)/9) (Njiti et al., 1996)

¶Location where disease ratings were recorded.

#Means of lines predicted to be homozygous resistant (Res) at the QTL locus on chromosome 1 based on the genetic markers Satt531, Satt179, and Satt468.

†† Means of lines predicted to be homozygous susceptible (Sus) at the QTL locus on chromosome 1 based on the genetic markers Satt531, Satt179, and Satt468.

‡‡Difference between the means of homozygous classes.

§§Probability value used to compare the mean of homozygous resistant and homozygous susceptible line.

References

Abdelmajid, K.M., L. Ramos, L. Leandro, G. Mbofung, D.L. Hyten, S.K. Kantartzi, R.L. Grier, V.N. Njiti, S. Cianzio, and K. Meksem. 2012. The 'PI 438489B'by 'Hamilton'SNP-based genetic linkage map of soybean [*Glycine max* (L.) Merr.] identified quantitative trait loci that underlie seedling SDS resistance. J. of Plant Genome Sci. 1:18-30.

Abney, T.S. and W.D. Crochet. 2006. Uniform soybean tests northern states, 2006. USDA-ARS, West Lafayette, IN.

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

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

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

Bao, Y., J.E. Kurle, G. Anderson, and N.D. Young. 2015. Association mapping and genomic prediction for resistance to sudden death syndrome in early maturing soybean germplasm. Mol. Breeding. 35:1-14.

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

Bell-Johnson, B., G. Garvey, J. Johnson, D. Lightfoot, and K. Meksem. 1998. Biotechnology approaches to improving resistance to SCN and SDS: Methods for high throughput marker assisted selection. Soybean Genet. Newsl. 25:115-117.

Cary, T.R. and B.W. Diers. 2004. Northern Regional soybean cyst nematode tests. University of Illinois, Urbana, IL.

Cooper, R.L., R.J. Martin, B.A. McBain, R.J. Fioritto, S.K. St. Martin, A. Calip-DuBois, and A.F. Schmitthenner. 1990. Registration of 'Ripley' soybean. Crop Sci. 30:963.

Cregan, P., T. Jarvik, A. Bush, R. Shoemaker, K. Lark, A. Kahler, N. Kaya, T.T. VanToai, D.G. Lohnes, J. Chung, and J.E. Specht. 1999. An integrated genetic linkage map of the soybean genome. Crop Sci. 39:1464-1490.

Cregan, P., and C. Quigley. 1997. Simple sequence repeat DNA marker analysis. p 173-185. In: Caetano-Anolles, G. and Gresshoff, P.M. (eds). DNA markers: Protocols, applications, and overviews, J. Wiley and Sons, New York.

de Farias Neto, A.L., G.L. Hartman, W.L. Pedersen, S. Li, G.A. Bollero, and B.W. Diers. 2006. Irrigation and inoculation treatments that increase the severity of soybean sudden death syndrome in the field. Crop Sci. 46:2547-2554.

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

Diers, B.W., T.R. Cary, D.J. Thomas, and C.D. Nickell. 2006. Registration of 'LD00-3309' soybean. Crop Sci. 46:1384.

Fan, J.B., A. Oliphant, R. Shen, B.G. Kermani, F. Garcia, K.L. Gunderson, M. Hansen. F. Steemers, S.L. Butler, P. Deloukas, L. Galver, S. Hunt, C. McBride, M. Bibikova, T. Rubano, J. Chen, E. Wickam, D. Doucet, W. Chang, D. Campbell, B. Zhang, S. Kruglyak, D. Bentley, J. Haas, P. Rigault, L. Zhou, J. Stuelpnagel, and M.S. Chee. 2003. Highly parallel SNP genotyping. Cold Spring Harb. Symp. Quant. Biol. 68:69–78.

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

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

Hartman, G., Y. Huang, R. Nelson, and G. Noel. (1997). Germplasm evaluation of glycine max for resistance to *Fusarium solani*, the causal organism of sudden death syndrome. Plant Dis. 81:515-518.

Hartman, G.L., L.F. Leandro, and J.C. Rupe. 2015. "Sudden Death Syndrome." In: G.L. Hartman, J.C. Rupe, E.J. Sikora, L.L. Domier, J.A. Davis, and K.L. Steffy (eds), Compendium of soybean diseases (5th ed). APS Press, St. Paul, MN.

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

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

Kandel, Y.R., C.A. Bradley, K.A. Wise, M.I. Chilvers, A.U. Tenuta, V.M. Davis, P.D. Esker, D.L. Smith, M.A. Licht, and D.S. Mueller. 2015. Effect of glyphosate application on sudden death syndrome of glyphosate-resistant soybean under field conditions. Plant Dis. 99:347-354.

Kao, C., and Z. Zeng. 1997. General formulas for obtaining the MLEs and the asymptotic variance- covariance matrix in mapping quantitative trait loci when using the EM algorithm. Biometrics. 53:653-665.

Kazi, S., J. Shultz, J. Afzal, J. Johnson, V. Njiti, and D.A. Lightfoot. 2008. Separate loci underlie resistance to root infection and leaf scorch during soybean sudden death syndrome. Theor. Appl. Genet. 116:967-977.

Keim, P., T. Olson, and R. Shoemaker. 1988. A rapid protocol for isolating soybean DNA. Soybean Genet. Newsletter-United States Department of Agriculture, Agricultural Research Service, 15.

Kim, E. 2007. Evaluation of soybean sudden death syndrome resistance and fungal growth on fungicide-amended medium. . M.S. thesis, Univ. of Illinois, Urbana.

Koenning, S.R., and J.A Wrather. 2010. Suppression of soybean yield potential in the continental United States from plant diseases from 2006 to 2009. Plant Health Prog. doi:10.1094/PHP-2010-1122-01-RS.

Littell, R.C., W.W. Stroup, G.A. Milliken, R.D. Wolfinger, and O. Schabenberger. 2006. SAS for mixed models. SAS institute.

Luckew, A., L. Leandro, M. Bhattacharyya, D. Nordman, D. Lightfoot, and S. Cianzio. 2013. Usefulness of 10 genomic regions in soybean associated with sudden death syndrome resistance. Theor. Appl. Genet. 126:2391-2403.

Mueller, D., G. Hartman, R. Nelson, and W. Pedersen. 2002. Evaluation of *Glycine max* germ plasm for resistance to *Fusarium solani* f. sp. glycines. Plant Dis. 86:741-746.

Mueller, D., R. Nelson, G. Hartman, and W. Pedersen. 2003. Response of commercially developed soybean cultivars and the ancestral soybean lines to *Fusarium solani* f. sp. *glycines*. Plant Dis. 87:827-831.

Nakajima, T., T. Mitsueda, and M.J.D. Charchar. 1993. Occurrence of soybean sudden-death syndrome caused by *Fusarium solani* in Brazil. Abstract. In 7th Int Fusarium Workshop, Pennsylvania State University, University Park. p (79).

Nickell, C.D., D.J. Thomas, T.R. Cary, D.E. Kyle, and J.M. Hegstad. 1998. Registration of 'Omaha' soybean. Crop Sci. 38:547.

Njiti, V., M. Shenaut, R. Suttner, M. Schmidt, and P. Gibson. 1996. Soybean response to sudden death syndrome: Inheritance influenced by cyst nematode resistance in Pyramid× Douglas progenies. Crop Sci. 36:1165-1170.

Njiti, V. N., R. J. Suttner, L. E. Gray, P. T. Gibson, and D. A. Lightfoot. 1997. Rate-Reducing Resistance to *Fusarium solani* f. sp. *phaseoli* underlies Field Resistance to Soybean Sudden Death Syndrome. Crop Sci. 37:132-138.

Njiti, V., T. Doubler, R.J. Suttner, L. Gray, P. Gibson, and D. Lightfoot. 1998. Resistance to soybean sudden death syndrome and root colonization by *Fusarium solani* f. sp. *glycines* in near-isogenic lines. Crop Sci. 38:472-477.

Njiti, V.N., M.A. Shenaut, R.J. Suttner, M.E. Schmidt, and P.T. Gibson. 1998. Relationship between soybean sudden death syndrome disease measures and yield components in F6-derived lines. Crop Sci. 38:673-678.

Njiti, V., K. Meksem, M. Iqbal, J. Johnson, M.A. Kassem, K. Zobrist, V.Y. Kilo, and D. Lightfoot. 2002. Common loci underlie field resistance to soybean sudden death syndrome in Forrest, Pyramid, Essex, and Douglas. Theor. Appl. Genet. 104:294-300.

Ploper, D. 1993. Sudden death syndrome: new soybean disease in northeastern Argentina. Agroindustrial Advances. 13:5-9.

Prabhu R.R., V.N. Nijiti, B. Bell-Johnson, J.E. Johnson, M.E. Schmidt, J.H. Klein, and D.A. Lightfoot. 1999. Selecting soybean cultivars for dual resistance to soybean cyst nematode and sudden death syndrome using two DNA markers. Crop Sci. 39:982-987.

Roy, K., D. Hershman, J. Rupe, and T. Abney. 1997. Sudden death syndrome of soybean. Plant Dis. 81:1100-1111.

Rupe, J.C., J.C. Correll, J.C. Guerber, C.M. Becton, E.E. Gbur, M.S. Cummings, and P.A. Yount. 2001. Differentiation of the sudden death syndrome pathogen of soybean, *Fusarium solani* f. sp. *glycines*, from other isolates of *F. solani* based on cultural morphology, pathogenicity, and mitochondrial DNA restriction fragment length polymorphisms. Can. J. Bot. 79:829-835.

SAS Institute. 2016. The SAS system for Microsoft Windows. Release 9.4. SAS Institute, Cary, NC.

Scherm, H., and X. Yang. 1996. Development of sudden death syndrome of soybean in relation to soil temperature and soil water matric potential. Phytopathology. 86:642-649.

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

Song, Q., G. Jia, Y. Zhu, D. Grant, R.T. Nelson, E. Hwang, D.L. Hyten, and P.B. Cregan. 2010. Abundance of SSR motifs and development of candidate polymorphic SSR markers (BARCSOYSSR_1. 0) in soybean. Crop Sci. 50:1950-1960.

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

SoyBase and the Soybean Breeder's Toolbox. Soybean breeder's toolbox genetic map information, available at <http://www.soybase.org>. Accessed January 30, 2017.

Srour, A., A.J. Afzal, L. Blahut-Beatty, N. Hemmati, D.H. Simmonds, W. Li, M. Liu, C. Town, H. Sharma, P. Arelli, and D.A. Lightfoot. 2012. The receptor like kinase at Rhg1-a/Rfs2 caused pleiotropic resistance to sudden death syndrome and soybean cyst nematode as a transgene by altering signaling responses. BMC Genomics. 13:368.

Triwitayakorn, K., V. Njiti, M. Iqbal, S. Yaegashi, C. Town, and D. Lightfoot. 2005. Genomic analysis of a region encompassing *QRfs1* and *QRfs2*: Genes that underlie soybean resistance to sudden death syndrome. Genome. 48:125-138.

Wang, D., J. Shi, S. Carlson, P. Cregan, R. Ward, and B. Diers. 2003. A low-cost, high-throughput polyacrylamide gel electrophoresis system for genotyping with microsatellite DNA markers. Crop Sci. 43:1828-1832.

Weems, J.D., J.S. Haudenshield, J.P. Bond, G.L. Hartman, K.A. Ames, and C.A. Bradley. 2015. Effect of fungicide seed treatments on *Fusarium virguliforme* infection of soybean and development of sudden death syndrome. Can. J. of Plant Pathol. 37:435-447.

Wen, Z., R. Tan, J. Yuan, C. Bales, W. Du, S. Zhang, M. Chilvers, C. Schmidt, Q. Song, P. Cregan, and D. Wang. 2014. Genome-wide association mapping of quantitative resistance to sudden death syndrome in soybean. BMC genomics. 15:809.

Wilcox, J.R., M.T. Roach, and T.S. Abney. 1989. Registration of 'Spencer' soybean. Crop Sci. 29: 830-831.

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

CHAPTER 5: STACKING OF ALLELES FROM MULTIPLE SOURCES INCREASES THE GENETIC RESISTANCE OF SOYBEAN TO HIGHLY VIRULENT SOYBEAN CYST NEMATODE ISOLATES*

Abstract

Soybean cyst nematode (SCN), *Heterodera glycines* (HG) Ichinohe, is estimated to be the pathogen that causes the greatest economic loss to soybean [*Glycine max* (L.) Merrill] in the USA. Genetic resistance is an effective way to manage SCN. Resistance sources have been identified, and resistance quantitative trait loci (QTL) from these sources have been mapped. However, there is a need to diversify SCN resistance genes in cultivars as most grown in the northern USA have resistance tracing only to the source PI88788. The objective of this study was to determine the effectiveness of combinations of SCN resistance alleles from different sources in two populations formed via backcrossing. Population 1 segregates for a resistance QTL from both PI567516C and PI88788 while Population 2 segregates for the same QTL as Population 1 and two QTL from PI468916. Lines from both populations were evaluated with two highly virulent nematode isolates that could overcome multiple sources of resistance. Furthermore, a subset of lines from Population 2 (Population 2 Subset) was evaluated with two additional nematode isolates. The SCN resistance alleles from each source significantly increased SCN resistance compared to the alternative alleles. The effect of resistance alleles varied depending on SCN isolate and population, and there was generally an increase in resistance as more resistance alleles were stacked together. These results indicate stacking multiple sources of resistance can be an effective means to increase broad-spectrum SCN resistance.

*Submitted to Crop Science. Brzostowski, L.F., and B.W. Diers. 2017. Crop Sci.

Introduction

Soybean cyst nematode (SCN) is the pathogen estimated to cause the largest economic loss to soybean in the USA. From 2006-2009, the estimated average annual yield loss to SCN was 3.5 million metric tons, which represents a loss of over a billion dollars a year for producers (Koenning and Wrather, 2010). The first identification of SCN in the USA was in North Carolina in 1953, and since this time, the pathogen has spread to most major soybean producing areas (Tylka and Marett, 2014).

Rotation to non-host crops such as corn can be a means of SCN control; however, this method does not completely eradicate the pathogen from the soil due to SCN's ability to overwinter in soil for several years (Porter et al., 2001; Jackson et al., 2005; Miller et al., 2006). Nematicide seed treatments have recently been made available to producers, but they need to be combined with genetic resistance to protect yields (Tylka et al., 2015; Tenuta and Tenuta, 2015). Genetic resistance is the most effective method to control losses due to SCN, and over the past several decades, SCN resistance has been a primary focus of soybean breeders (Kopisch-Obuch et al., 2005). The mechanism of SCN resistance is not entirely understood. In resistant cultivars, the nematode penetrates the root and forms a syncytium; however, the syncytium either forms slowly or becomes necrotic soon after formation leading the nematode to starve to death (Williamson and Hussey, 1996).

Genetic resistance to SCN in soybean was identified shortly after the pathogen was discovered in the USA. Caldwell et al. (1960) observed resistance from PI548402 ('Peking') was controlled by three genes named *rhg1*, *rhg2*, and *rhg3*. In a later study, Matson and Williams (1965) reported a fourth resistance gene from Peking named *Rhg4*. Rao-Arelli et al. (1992) identified another gene in PI88788, which was given the designation *Rhg5* (Rao-Arelli et al., 1992). SCN resistance was shown to be inherited as a quantitative trait in genetic mapping studies, and resistance genes were mapped as quantitative trait loci (QTL) (Concibido et al., 2004). Two major mapped SCN resistance QTL from Peking were given the gene designations *rhg1* and *Rhg4*, and these designations will be used in this article. The terms gene and QTL will be used interchangeably to simplify explanations.

Over 118 soybean accessions have been identified as resistant to SCN and therefore are considered potential resistance sources (Arelli et al., 2000). Despite the number of resistance

sources available, resistance in cultivars has a narrow base with only seven resistance sources commonly used by breeders. Several of these sources were found to share one or more resistance genes, further limiting diversity for resistance. In the northern USA, PI88788 is the most widely used SCN resistance source in commercial cultivars. It has resistance to a broad range of HG (*Heterodera glycines*) types (Niblack et al., 2002), and cultivars with good agronomic traits have been developed from this source (Diers and Arelli, 1999; Cary and Diers, 2015). In the 2016 University of Illinois Soybean Variety Test, 254 SCN resistant cultivars entered in the test had their source of SCN resistance listed. Of these cultivars, 249 (98%) had their resistance from PI 88788 (University of Illinois, 2016).The resistance allele at *rhg1* from PI88788 was shown to be different from the allele at this locus from Peking, and the PI88788 allele was given the designation, *rhg1-b* (Kim et al., 2010). *Rhg1* has been cloned, and the resistance has been determined to be conferred by copy number variation of three genes: an amino acid transporter, an α-SNAP protein, and a WI12 (wound-inducible domain) protein (Cook et al., 2012).

With the heavy reliance on just a few sources of resistance, it is crucial to continue to search for and evaluate novel resistance sources to protect yields. This need is even more important when the ability of field nematodes to adapt to and overcome host plant resistance is taken into account. In a survey of soil samples taken around Illinois, 70% of the SCN populations studied could overcome resistance conferred by PI88788 (Niblack et al., 2008).

Young (1999) identified PI567516C as the only source of resistance to LY1, a highly virulent synthetic population of nematodes formed by mass mating HG type 1.2.3 females to HG type 1.2 males. Follow-up research confirmed this resistance in PI567516C and its genetic distinction from 'Hartwig' (Anand, 1992), a highly resistant cultivar with resistance from PI437654 (Chen et al., 2006; Arelli et al., 2009). In a population derived from a cross between PI567516C and Hartwig, a genetic region on chr 10 (formally linkage group (LG) O), defined by the markers Satt592, Satt331, and Sat_274, was associated with resistance to LY1 (Arelli et al., 2010). A QTL was mapped to the same region on chr 10 in a population developed from a cross between PI567516C and the susceptible cultivar 'Magellan' (Schapaugh et al., 1998; Vuong et al., 2010). The QTL conferred resistance to HG types 2.5.7, 0, 2.7, 1.3.5.6.7, and LY1, and it was confirmed in a recombinant inbred line (RIL) population derived from the original PI567516C by Magellan cross.

Breeders have also sought to increase the genetic diversity of SCN resistance by identifying resistant soybean relatives and using them as parents in mapping studies. Wang et al. (2001) mapped QTL alleles that significantly increased SCN resistance from the *Glycine soja* Siebold & Zucc. accession PI468916 on both chr 15 (LG E) and chr 18 (LG G). The QTL on chr 15 and chr 18 were later confirmed (Kabelka et al., 2005). The chr 15 QTL was designated cqSCN-006, and the chr 18 QTL was designated cqSCN-007. Further testing of these QTL provided no evidence of linkage drag that would reduce yield (Kabelka et al., 2006). Kim and Diers (2013) fine mapped cqSCN-006 to a 803.4 kb region and cqSCN-007 to a 146.5 kb region. The region containing cqSCN-006 was recently reduced to a 212.1 kb interval and cqSCN-007 to a 103.2 kb interval (Yu and Diers, 2016). The abundance of genetic and phenotypic information for these QTL make them good candidates for incorporation into other genetic backgrounds to determine whether they are appropriate for widespread use in breeding programs.

Stacking multiple QTL from different resistance sources may provide more durable protection against SCN isolates that can overcome the resistance conferred by the genes commonly used by breeders. The objectives of this study are to i) test the effect of SCN on a population segregating for *rhg1-b* from PI88788 and the resistance allele from PI567516C for the chr 10 QTL (herein referred to as CHR10) ii) test the effect of SCN on a population segregating for *rhg1-b* from PI88788 and the resistance alleles at cqSCN-006, cqSCN-007, and CHR10.

Materials and Methods

Population Development

Population 1

Population 1 segregated for *rhg1-b* from PI88788 and the SCN resistance QTL CHR10 from PI567516C. Population development was initiated by crossing the recurrent parent LD00-3309 (Diers et al., 2006), a maturity group IV cultivar with SCN resistance tracing back to PI88788, with PI567516C, an accession which does not have a resistance allele at *rhg1* but does have a resistance allele at CHR10. Following each generation of crossing, plants were genotyped

with markers linked to the QTL. Two generations of backcrossing were conducted to reach the BC_2F_1 generation. BC_2F_1 plants fixed for the *rhg1-b* allele and heterozygous for CHR10 were crossed to LD09-15628, a line developed by backcrossing the susceptible allele at *rhg1* into the background of LD00-3309. The cultivar IA3023 was the donor parent of the susceptible allele, and the line had been developed through four backcrosses (BC_4). The F_1 plant from the cross between LD09-15628 and the BC_2F_1 plant previously described were heterozygous for *rhg1-b* and CHR10. The population was advanced through single-seed descent in a greenhouse and field until the F_4 generation. After testing each F_4 plant with markers linked to the QTL, heterozygous plants were eliminated, and each plant was hand harvested to produce 143 $F_{4:5}$ lines (Table A.1).

Population 2

Population 2 segregated for *rhg1-b* from PI88788, the resistance QTL cqSCN-006 and cqSCN-007 from PI468916, and the resistance QTL CHR10 from PI567516C. This population was developed by first crossing the recurrent parent 09SCNPOP11-9 to PI567516C. 09SCNPOP11-9 is a $BC_4F_{2:3}$ line fixed for the resistance alleles at cqSCN-006 and cqSCN-007 and for *rhg1-b* in the LD00-3309 background. After each generation of crossing, marker-assisted selection was performed with markers linked to the QTL. Two generations of backcrossing were conducted, and BC_2F_1 plants fixed for the *rhg1-b* allele and heterozygous for cqSCN-006, cqSCN-007, and CHR10 were crossed to LD09-15628. An F_1 plant from this cross that was predicted to be heterozygous for *rhg1-b*, cqSCN-006, cqSCN-007, and CHR10 was selected with genetic markers. Population 2 was advanced by single-seed descent in the greenhouse and field until the F_4 generation. Each F_4 plant was genotyped with markers linked to the QTL, and heterozygous plants were eliminated. A total of 107 F_4 plants were grown to maturity and individually harvested to produce $F_{4:5}$ lines (Table A.2).

Genetic Analysis

The F_4 plants in Populations 1 and 2 were genotyped to predict the resistance allele for each plant and to identify those plants homozygous for genetic markers linked to the QTL. Genomic DNA was isolated from young trifoliolate leaves via a modified CTAB (cetyltrimethyl

ammonium bromide) method (Keim et al., 1988). The DNA samples were tested with simple sequence repeat (SSR) markers linked to the QTL of interest (Cregan et al., 1999; Vuong et al., 2010; Kim and Diers, 2013) by conducting polymerase chain reactions (PCR) according to the method described by Cregan and Quigley (1997). Products from PCR were separated on 6% (w/v) nondenaturing polyacrylamide gels by electrophoresis (Wang et al., 2003).

SCN Bioassay

SCN bioassays for Populations 1 and 2 were conducted using two different SCN field isolates; the first was a HG type 1.2.3.4.5.6.7 (TN21) and the second a HG type 1.2.3.5.6.7. The bioassays were conducted separately for each isolate by infesting a single F_6 plant from each F_4-derived line. The SCN isolates had been maintained for several generations and were obtained from Alison Colgrove at the University of Illinois, Urbana, Illinois. TN21 was originally collected from a field in Missouri where it was selected on PI437654 and then Hartwig, while HG 1.2.3.5.6.7 was collected from a field in DeKalb, IL.

A subset of thirteen F_4-derived lines from Population 2 (Population 2 Subset) was evaluated with two additional isolates. Lines were replicated three times by infesting F_6 plants with SCN isolates HG 2.5.7 and HG 1.2.3.4.5.6.7 (TN23). Tests were conducted separately according to isolate. HG 2.5.7 was collected in a field in southwest White County, IL. TN23 was collected in a field at the former Dixon Springs Agricultural Center in Simpson, IL, and it was originally selected on PI437654.

In all tests, the following check cultivars and indicator lines were replicated four times: Peking, PI88788, PI90763, PI437654, PI209332, PI89722, ‘Cloud’, ‘Lee 74’, IA3023, LD00-3309, PI567516C, and LD00-2817 (Diers et al., 2010). Lines, checks, and indicator lines were planted in a randomized design.

The bioassays were conducted in a greenhouse in a thermo-regulated water bath system using modified methods described by Arelli et al. (2000) and Niblack et al. (2002). Briefly, PVC (polyvinyl chloride) tubes filled with a 1:1 sand to soil mix were placed in a plastic crock. Each tube was inoculated with approximately 2,000 nematode eggs suspended in water. A single germinated seedling was planted in each tube and an experimental unit was a tube with a plant. The crocks were suspended in a water bath maintained at 27±1°C. The plants were watered as

needed and grown under a 16-hour day length. After the test was maintained for 28 days, the tubes are immersed in water to remove the plants from the soil. Cysts were dislodged from each root on a nested 850-µm aperture over 250-µm aperture sieve using gentle water pressure, and the cysts were counted using a stereo microscope. A female index (FI) was calculated on a per plant basis using the following equation (Golden et al., 1970): FI = 100 x (Number of cysts per plant/Average number of cysts on susceptible host "Lee 74"). Lines were then classified according to Schmitt and Shannon (1992): resistant (FI<10), moderately resistant (FI=10-30), moderately susceptible (FI=31-60), and susceptible (FI>60).

Statistical Analysis

Marker-trait associations were tested individually using (ANOVA) by the PROC GLM function in SAS v9.4 (SAS Institute Inc., Cary, NC). Markers individually significant at α=0.05 were then placed in multivariate models, and all two-way and three-way interactions between markers were evaluated. Those not significant were removed from the model and analysis repeated. All factors were considered fixed. If the residuals were not normal with homogeneous variances, female indices were $\log_{10}$ transformed. Reported p values are based on transformed data; however, all other data have been back transformed to original units. R^2 values from the multivariate ANOVA measure the percentage of total genotypic variance for resistance explained by the QTL and their interactions. Means of genotypic classes were separated according to Fisher's LSD.

Results

HG Type Tests

The SCN isolates, HG 1.2.3.5.6.7, TN21, HG 2.5.7, and TN23, are thought to be near-homogenous because they have been maintained in the greenhouse for several generations. Female reproduction of all isolates on the susceptible check Lee 74 was robust with a FI range of 223 for TN23 to 402 for HG 1.2.3.5.6.7 (Table 5.1). All four isolates reproduced as expected on the indicator lines.

Population 1

Population 1 segregated for resistance alleles from PI88788 and PI567516C in a LD00-3309 background. The FI of lines in Population 1 had a continuous distribution for both SCN isolates (Fig. A.1A-B). The range of FI for lines was 30 to 203 for HG 1.2.3.5.6.7 and 12 to 118 for TN21. LD00-3309, the recurrent parent and donor of the *rhg1-b* allele, had a FI of 65 for HG 1.2.3.5.6.7 and 42 for TN21. Phenotypic data were collected on 140 lines in the HG 1.2.3.5.6.7 test and 141 lines in the TN21 test.

Across all lines within the HG 1.2.3.5.6.7 and TN21 tests, significant ($P<0.0001$) QTL effects were observed for *rhg1* and CHR10 (Table 5.2) resulting in lines containing the resistance alleles being more resistant than those with the alternative alleles. The magnitude of the effects was dependent on the QTL and SCN isolate. For HG 1.2.3.5.6.7, FI average across all lines in the population was 90. The difference between the homozygous classes for *rhg1* was a FI of -38, and this difference for CHR10 was -26. These negative values represent a decrease in reproduction due to the resistance alleles. There was no significant interaction between *rhg1* and CHR10, and the two QTL together explained 40% of the total variation for resistance to HG 1.2.3.5.6.7. For TN21, the average FI across all lines was 52, and the difference between the homozygous classes was -19 for *rhg1* and -15 for CHR10. There was a significant interaction between the two loci for TN21, and *rhg1*, CHR10, and the interaction explained 50% of the total variation for resistance.

When genotypic classes were compared, similar trends were observed in the HG 1.2.3.5.6.7 and TN21 tests (Fig. 5.1A-B). Lines that were homozygous resistant at one or two loci had a reduced FI compared to lines containing no resistance QTL. Lines with the resistance allele at either CHR10 or *rhg1* were not significantly different from each other. Additionally, lines with the resistance alleles at both QTL had a reduced average FI compared to lines with only one resistance allele or lines that were homozygous susceptible. Although adding resistance alleles resulted in improved resistance, the significant statistical interaction between *rhg1* and CHR10 observed in the TN21 test was the result of the combined effect of the resistance QTL not being additive. This means that the increase in resistance was smaller than expected when the two resistance QTL were combined based on effect of the individual resistance alleles. For both isolates, combined effect of the resistance alleles for both QTL reduced the FI by approximately

50%. For TN21, the average FI of lines with resistance alleles for both QTL was 38 while lines with no resistance alleles averaged 72. For HG 1.2.3.5.6.7, the mean FI for those lines with both resistance alleles was 62 while the average FI for lines with neither resistance allele was 125.

Population 2

Population 2 segregated for resistance alleles from PI88788, PI468916, and PI567516C in an LD00-3309 background. There was a continuous distribution in the FI of lines in Population 2 for both SCN isolates (Fig. A.1C-D). The range of FI for lines was 1 to 167 for HG 1.2.3.5.6.7 and 11 to 116 for TN21. Phenotypic data were collected on 105 lines in the HG 1.2.3.5.6.7 test and 104 lines in the TN21 test.

Across all lines, each of the four resistance alleles significantly ($P<0.05$) increased resistance to both HG 1.2.3.5.6.7 and TN21, and the magnitude of the QTL effect was dependent upon the QTL and SCN isolate (Table 5.2). For HG 1.2.3.5.6.7, the average FI of the population was 33. cqSCN-006 had a numerically greater effect on reproduction than the other QTL, and the FI difference between the homozygous classes ranged from -15 to -30 for the four QTL. A statistically significant interaction ($P<0.01$) was observed between *rhg1* and cqSCN-007. The combined effect of the QTL and interaction explained 57% of the total variance for resistance to HG 1.2.3.5.6.7. For TN21, the average FI of the population was 48, and *rhg1-b* had a numerically greater effect on reproduction than the other resistance alleles. The differences in FI between the homozygous QTL classes ranged from -7 to -18. There were no significant interactions between QTL, and together the QTL explained 38% of total genetic variation associated with resistance to TN21.

For HG 1.2.3.5.6.7 and TN21, the mean FI of lines containing the resistance alleles at two or more QTL were significantly ($P<0.05$) lower than the mean of lines with no resistance alleles (Fig. 5.2A-B). Two exceptions to this were the means of lines with resistance alleles at *rhg1* and CHR10 in the HG 1.2.3.5.6.7 test and lines with the cqSCN-006 and CHR10 resistance alleles in the TN21 test. Generally, lines homozygous for all four resistance QTL alleles had reduced SCN reproduction compared to lines that only had a single QTL or lines that were homozygous susceptible. Lines carrying all four resistance QTL and lines carrying resistance alleles at *rhg1,* cqSCN-007, and CHR 10 were classified as resistant to HG 1.2.3.5.6.7 and had

average FIs of 4 and 7, respectively (Fig. 5.2A). Lines with the four resistance alleles and lines with the resistance alleles *rhg1-b,* cqSCN-006, and CHR10 were classified as moderately resistant to TN21 with average FIs of 29 and 28, respectively (Fig. 5.2B).

Population 2 Subset

Population 2 Subset consisted of the following genotypes: three lines containing no resistance alleles; three lines containing only the *rhg1-b* resistance allele; three lines containing the *rhg1-b* and CHR10 resistance alleles; and four lines containing the *rhg1-b*, cqSCN-006, cqSCN-007, and CHR10 resistance alleles. With the resistance QTL detected in Population 1 and Population 2, the Population 2 Subset allowed for more extensive testing of resistance QTL stacks with two additional SCN isolates, HG 2.5.7 and TN23. There was a continuous distribution for the average FI of individual lines in the Population 2 Subset for both SCN isolates, and the FI of lines ranged from 10 to 88 for HG 2.5.7 and 23 to 100 for TN23 (Fig. A.1E-F). The average female index on LD00-3309 for HG 2.5.7 was 59 and was 37 for TN23.

For HG 2.5.7, lines that were homozygous for the resistance allele at all four QTL had a reduced average FI compared to the other genotypic classes (Fig. 5.3A). The average FI of these lines was 15 which gives them a moderately resistant classification. Lines homozygous for the four susceptibility alleles had an average FI of 71, and this FI was decreased to 49 through the addition of *rhg1-b*. The FI of lines homozygous for both CHR 10 and *rhg1-b* was 59, which was not significantly different from lines carrying *rhg1-b* alone. An improvement in resistance occurred when the resistance alleles at cqSCN-006 and cqSCN-007 were stacked with the *rhg1-b* and CHR10 resistance alleles.

For TN23, lines with the two or four resistance QTL stacks had a significantly reduced female index compared to the other genotypic classes, and they were not significantly different from each other (Fig. 5.3B). The average FI of lines with all four resistance QTL alleles was 28 while the average of lines with the *rhg1-b* and CHR10 resistance allele stack was 33. Lines carrying the *rhg1-b* allele had a reduced FI compared to lines with no resistance QTL alleles but had an increased FI compared to lines with resistance stacks. Lines with no resistance QTL alleles were susceptible with an average FI of 79.

Discussion

The SCN isolates in this study were selected for their ability to overcome several resistance sources that are commonly used by breeders. These isolates were naturally occurring; however, highly virulent SCN populations such as HG 1.2.3.5.6.7, TN21, and TN23 are currently rare in field environments. The widespread dependence on PI88788 and a handful of other resistance sources increases the likelihood that more nematode field populations will overcome the resistance conferred by these sources. SCN populations that can reproduce on PI88788 such as HG 2.5.7 are already common in soybean fields across the USA and Canada (Niblack et al., 2008; Faghihi et al., 2010; Brzostowski et al., 2014). In a survey of 527 soil samples collected across Tennessee, Indiana, Illinois, and Ontario, >61% of the isolates collected had FI >10% on PI88788 and >56% of the isolates collected within a state had a FI>10 on at least two differentials (Faghihi et al., 2010). This distribution of nematode isolates demonstrates the need to develop cultivars with more broad and durable resistance to SCN.

The soybean populations used in this study were developed to test the effect of combinations of the resistance alleles *rhg1-b* from PI88788, cqSCN-006 and cqSCN-007 from PI468916, and CHR10 from PI567516C on SCN reproduction. These QTL were incorporated into an LD00-3309 background to form segregating populations of isolines. The QTL are known to have large effects; therefore, population sizes were relatively small. The LD00-3309 background was selected because it is representative of high-yielding cultivars with PI88788 resistance that are adapted to the midwestern USA (Diers et al., 2006). All four resistance QTL were detected across multiple nematode isolates with each conferring partial resistance, and the magnitude of their effect on FI varied according to QTL and isolate. When lines contained multiple resistance alleles were considered, lines classified as resistant and moderately resistant to multiple nematode isolates were observed. CHR10 from PI567516C has been noted as a promising novel SCN resistance QTL because it has been shown to confer resistance to several SCN isolates (Young, 1999; Vuong et al., 2010). In our study, CHR10 generally had a smaller effect on nematode reproduction compared to the other QTL tested; however, the addition of the CHR10 resistance allele did increase resistance in several combinations that it was evaluated most notably in Population 1 (Fig. 5.1). Overall, our data suggest broad spectrum resistance to SCN can be significantly improved by stacking resistance alleles from multiple sources.

The accession PI437654 confers resistance to a broad range of SCN isolates and has been used to develop highly resistant cultivars and germplasm lines (Anand et al., 1988; Diers et al., 1997; Wu et al., 2009; Diers et al. 2010). Despite its promise, few cultivars with resistance from PI437654 have been released because it has been difficult to develop agronomically acceptable high yielding cultivars with this source. Additionally, resistance conferred by PI437654 is mediated by numerous QTL, which makes it challenging to achieve resistance similar to PI437654 in elite cultivars (Concibido et al., 2004; Wu et al., 2009). For example, we observed that when inoculated with HG 1.2.3.5.6.7, LD00-2817, a high yielding cultivar with SCN resistance from PI437654 and/or Peking that was bred from Hartwig had a FI of 58, but PI437654 only had a FI of 1 (Table 5.1). Despite the high level of resistance provided by PI437654, two of the isolates used in our study had a FI>10 on PI437654, TN21 had an average FI of 44 while TN23 had an average FI of 80. A reason for the high reproduction on PI437654 is that both isolates were selected on PI437654 or Hartwig, which has its resistance tracing to PI437654. Our lines with the four gene stacks had numerically reduced FI compared to PI437654 with a FI of 29 for TN21 and 28 for TN23 (Table 5.1; Figs. 5.1-3).

Pyramiding resistance in a single genetic background has been demonstrated as an effective way to improve the host resistance to soybean pathogens and pests, including SCN. Kim et al. (2011) stacked SCN resistance alleles from exotic and domestic sources and observed increased resistance to the SCN isolates PA1, PA3, and PA5. Qualitative or quantitative gene pyramids in soybean have also been shown to enhance resistance to *Soybean Mosaic Virus* (qualitative), soybean aphid (qualitative), Phytopthora root rot (quantitative), and leaf chewing insects (quantitative) (Dorrance et al., 2008; Shi et al., 2009; Li et al., 2010; Ajayi-Oyetunde et al., 2016; Ortega et al., 2016).

It will be important to evaluate the effect of stacking resistance QTL on agronomic traits including yield. One of the reasons for the widespread use of PI88788 as a resistance source is that it has been combined with high yield and other good agronomic traits. In some genetic backgrounds, *rhg1* has been associated with yield drag in environments with low SCN pressure; however, there is evidence that the genetic linkage between *rhg1* and yield depression can be broken (Mudge et al. 1996, Kopisch-Obuch et al., 2005). Kabelka et al. (2006) evaluated the yield of one population segregating for resistance at cqSCN-006 and cqSCN-007 and another segregating for resistance at cqSCN-006, cqSCN-007, and *rhg1* from PI88788 in environments

with varying SCN pressure. Overall, the resistance alleles did not have a negative effect on yield and in some cases enhanced yield. For cultivars with stacked resistance sources to be truly effective, economic yields must be maintained.

Phenotyping for SCN resistance is costly and laborious; therefore, marker-assisted selection (MAS) for SCN resistance is of great interest to breeders (Concibido et al., 2004). While SSR markers linked to SCN resistance QTL were used in this study and have been used in breeding for almost two decades, rapid advances and decreasing costs in sequencing technologies have led to the recent development of low cost, high throughput single nucleotide polymorphism (SNP) marker-based assays for SCN resistance (Kadam et al., 2016). These SNP assays are not only more efficient with a lower error rate than previous genotyping methods, they can also identify the copy number variation at the *rhg1* locus.

It is expected that the next generation of breeding will lead to improved SCN resistance. High-throughput genotyping platforms could identify new alleles and genes that can be stacked using MAS to create cultivars with broad-based resistance to SCN. The combination of predictive modeling with SNP data could also reduce the time needed to stack multiple resistance QTL into elite cultivars and to select for these QTL in breeding programs (Bao et al., 2014; Shi et al., 2015). Population 1 and Population 2 took several years to develop and while pyramiding resistance is often suggested as a means to improve resistance to pests and pathogens, often the time needed discourages breeders from doing so.

Current genetic resistance to SCN is narrow, and breeders must implement new strategies to effectively manage this pathogen. The objective of our study was to evaluate the effect of stacking novel and common SCN resistance alleles on the reproduction of nematode isolates that can overcome multiple resistance sources. These isolates could not be controlled by *rhg1-b* mediated resistance alone. Based on our results, PI468916 and PI567516C are alternative sources breeders can use to enhance and diversify SCN resistance in their programs. Combining resistance alleles from PI88788, PI468916, and PI567516C conferred resistance or partial resistance to highly virulent nematode isolates. This suggests stacking resistance sources improves resistance to SCN and can be a useful strategy for future breeding. Genetic resistance remains important to controlling losses to SCN. Incorporating new resistance sources into breeding programs and stacking resistance sources will provide the necessary options needed for producers to protect their yields from SCN damage.

Acknowledgements

This research was supported by funding to LFB and BWD from the United Soybean Board and by funding from the North Central Soybean Research program to BWD.

Tables and Figures

Table 5.1. The reaction of four soybean cyst nematode (SCN) isolates to Lee 74, the *Heterodera glycines* type indicator lines, and elite checks with known genetic resistance to SCN.

			Female Index on Indicator Lines and Elite Checks†								
Isolate	# Females on Lee 74	HG type‡	1§ Peking	2 PI88788	3 PI90763	4 PI437654	5 PI209332	6 PI89772	7 Cloud	LD00-3309	LD00-2817
1.2.3.5.6.7	402	1.2.3.5.6.7	45	13	11	1	15	15	53	65	58
TN21	371	1.2.3.4.5.6.7	66	23	45	44	26	34	43	42	44
HG 2.5.7	303	2.5.7	2	31	0	1	22	0	39	59	6
TN23	223	1.2.3.4.5.6.7	72	23	70	80	21	58	78	37	83

†Female index=100 x (average number of females per line/average number of females on 'Lee 74') (Golden et al., 1970).

‡HG Type of isolate based on female indices.

§ Numerical designation of indicator line in HG Type test (Niblack et al., 2002).

Table 5.2. Effects of soybean cyst nematode (SCN) resistance quantitative trait loci (QTL) on female reproduction in Populations 1 and 2 for HG 1.2.3.5.6.7 and TN21 nematode isolates.

Population	QTL	HG 1.2.3.5.6.7 Female Indices					HG TN21 Female Indices				
		Resistant allele†	Susceptible allele‡	Effect§	P value	R^2¶	Resistant allele	Susceptible allele	Effect	P value	R^2
Population 1	rhg1	70	108	-38	<.0001	0.27	42	61	-19	<.0001	0.31
	CHR10	76	102	-26	<.0001	0.12	44	59	-15	<.0001	0.18
Population 2	rhg1	22	43	-21	<.0001	0.09	38	56	-18	<.0001	0.16
	cqSCN-006	21	43	-22	<.0001	0.06	39	56	-17	<.0001	0.12
	cqSCN-007	17	47	-30	<.0001	0.24	43	51	-8	0.0120	0.01
	CHR10	25	40	-15	<.0001	0.07	44	51	-7	0.0315	0.04

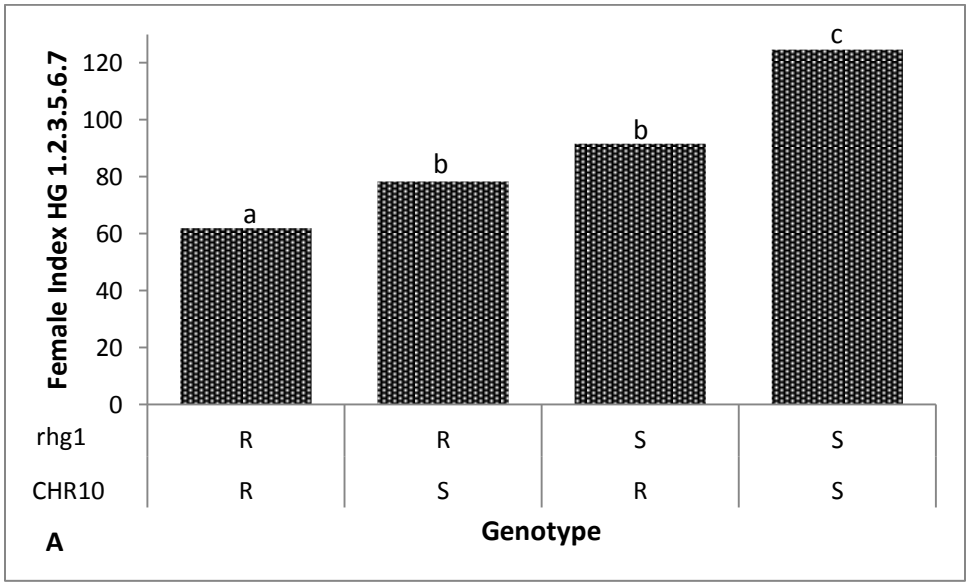
†Mean of lines homozygous for the SCN resistance allele at the locus.

‡Mean of lines homozygous for the SCN susceptible allele at the locus.

§Difference between homozygous classes.

¶Proportion of total genetic variation for resistance explained by the QTL.

Figure 5.1. Average female index values for Population 1 genotypes when evaluated with HG 1.2.3.5.6.7 (A) and TN21 (B) nematode isolates. R indicates homozygous for the resistance allele of the QTL while S indicates homozygous susceptible. Means with the same letter are not significantly different (α=0.05). Mean separations are based on $\log_{10}$ transformed data.



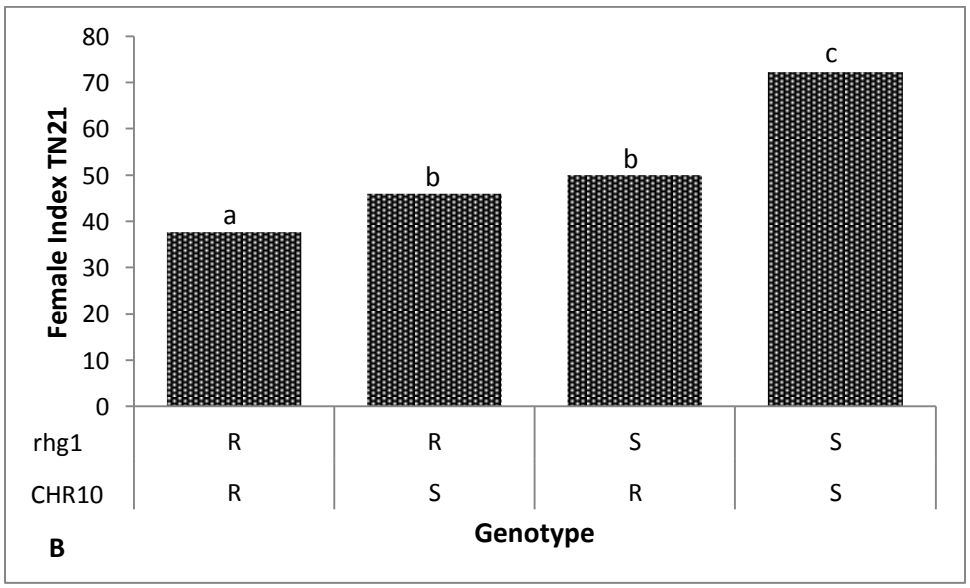
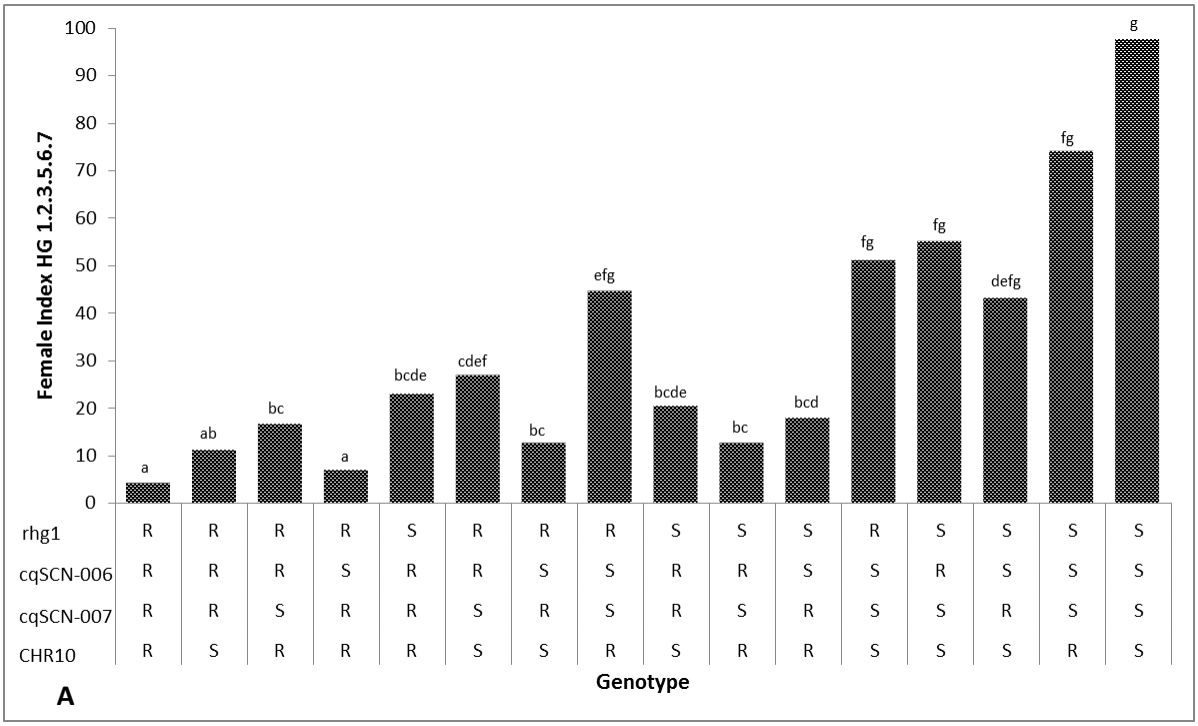


Figure 5.2. Average female index values for Population 2 genotypes when evaluated with HG 1.2.3.5.6.7 (A) and TN21 (B) nematode isolates. R indicates homozygous for the resistance allele of the QTL while S indicates homozygous susceptible for the QTL. Means with the same letter are not significantly different (α=0.05). For the HG 1.2.3.5.6.7 isolate data, mean separations are based on $\log_{10}$ transformed data.



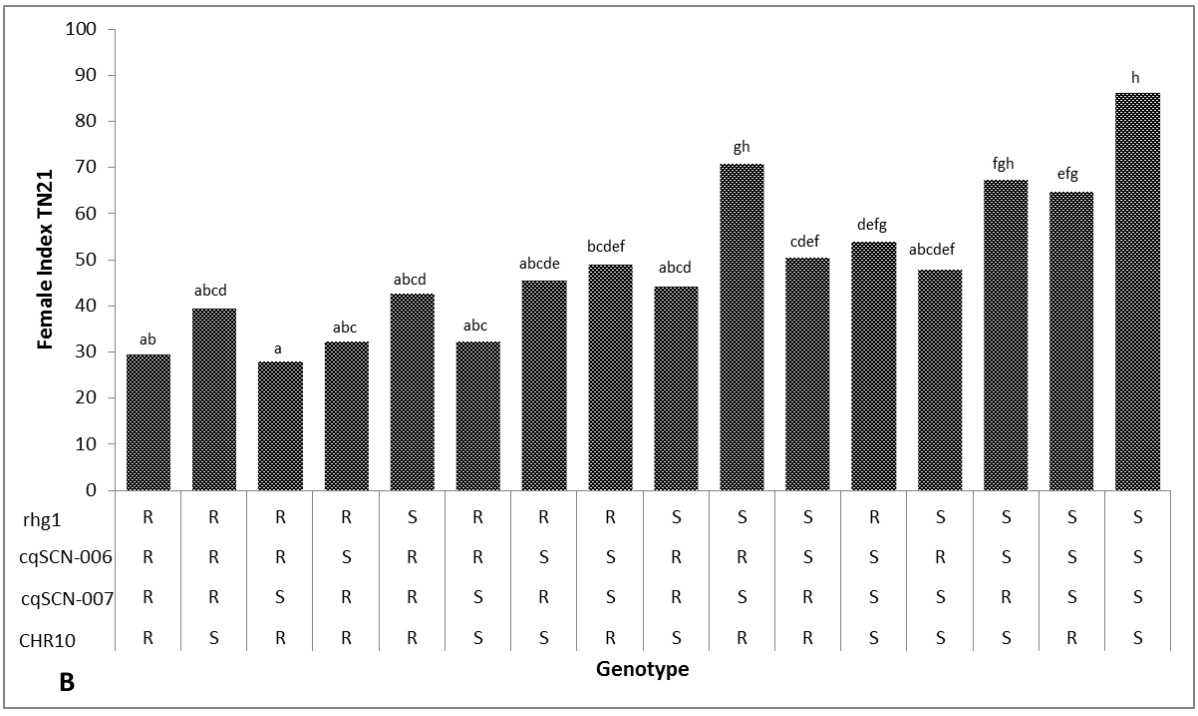
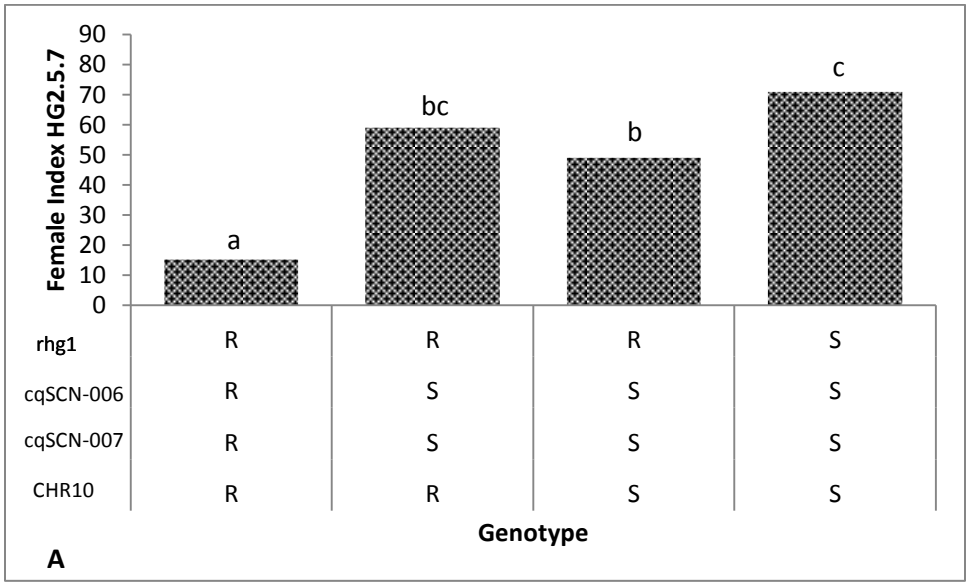
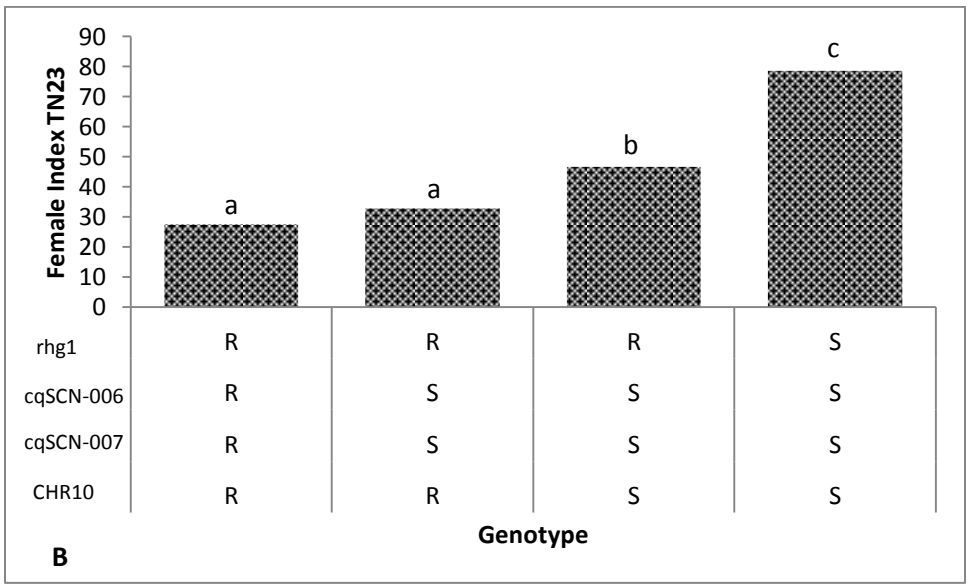


Figure 5.3. Average female index values for Population 2 Subset genotypes when evaluated with HG 2.5.7 (A) and TN23 (B) nematode isolates. R indicates homozygous for the resistance allele of the quantitative trait loci (QTL) while S indicates homozygous susceptible at the QTL. Means with the same letter are not significantly different (α=0.05). Mean separations are based on $\log_{10}$ transformed data.





References

Anand, S.C. 1992. Registration of 'Hartwig' soybean. Crop Sci. 32:1069-1070.

Anand, S.C., K.M. Gallo, I.A. Baker, and E.E. Hartwig. 1988. Soybean plant introductions with resistance to races 4 or 5 of soybean cyst nematode. Crop Sci. 28:563-564.

Arelli, P.R., V.C. Concibido, and L.D. Young. 2010. QTLs associated with resistance in soybean PI567516C to synthetic nematode population infecting cv. Hartwig. J. of Crop Sci. and Biotech. 13:163-167.

Arelli, P.R., D.A Sleper, P. Yue, and J.A. Wilcox. 2000. Soybean reaction to races 1 and 2 of *Heterodera glycines*. Crop Sci. 40:824-826.

Arelli, P.R., L.D Young, and V.C. Concibido. 2009. Inheritance of resistance in soybean PI 567516C to LY1 nematode population infecting cv. Hartwig. Euphytica. 165:1-4.

Ajayi-Oyetunde, O.O., B.W. Diers, D. Lagos-Kutz, C.B. Hill, G.L. Hartman, U. Reuter-Carlson, and C.A. Bradley. 2016. Differential reactions of soybean isolines with combinations of aphid resistance genes *Rag1*, *Rag2*, and *Rag3* to four soybean aphid biotypes. J. Econ. Entomol. 109:1-7.

Bao, Y., T. Vuong, C. Meinhardt, P. Tiffin, R. Denny, S. Chen, H.T. Nguyen, J.H. Orf, and N.D. Young. 2014. Potential of association mapping and genomic selection to explore PI 88788 derived soybean cyst nematode resistance. The Plant Genome. 7:3

Brzostowski, L.F., W.T. Schapaugh, P.A. Rzodkiewicz, T.C Todd, and C.R. Little. 2014. Effect of host resistance to *Fusarium virguliforme* and *Heterodera glycines* on sudden death syndrome disease severity and soybean yield. Plant Health Prog. doi:10.1094/PHP-RS-13-0100.

Caldwell, B.E., C. Brim, and J. Ross. 1960. Inheritance of resistance of soybeans to the cyst nematode, *Heterodera glycines*. Agron. J. 52:635-636.

Cary, T.R., and B.W. Diers. 2015. Northern regional soybean cyst nematode tests. University of Illinois, Urbana, IL.

Chen, Y., D. Wang, P. Arelli, M. Ebrahimi, and R.L. Nelson. 2006. Molecular marker diversity of SCN-resistant sources in soybean. Genome. 49:938-949.

Cook, D.E., T.G. Lee, X. Guo, S. Melito, K. Wang, A.M. Bayless, J. Wang, T.J. Hughes, T.E. Clemente, and B.W. Diers. 2012. Copy number variation of multiple genes at *Rhg1* mediates nematode resistance in soybean. Science. 338:1206-1209.

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

Cregan, P.B., J. Mudge, E.W. Fickus, D. Danesh, R. Denny, and N.D. Young. 1999. Two simple sequence repeat markers to select for soybean cyst nematode resistance conditioned by the *rhg1* locus. Theor. Appl. Genet. 99:811-818.

Cregan, P., and C. Quigley. 1997. Simple sequence repeat DNA marker analysis. In: G. Caetano-Anolles and Gresshoff, P.M. (eds), DNA markers: Protocols, applications. and overviews, John Wiley & Sons, New York. p. 173-185.

Diers, B.W., and P.R. Arelli. 1999. Management of parasitic nematodes of soybean through genetic resistance. In: Kauffman, H. E. (ed), Proc. of World Soybean Research Conference VI. Superior Printing, Champaign, IL. p. 300-306.

Diers, B.W., T. Cary, D. Thomas, A. Colgrove, and T. Niblack. 2010. Registration of LD00-2817P soybean germplasm line with resistance to soybean cyst nematode from PI 437654. J. of Plant Registrations. 4:141-144.

Diers, B.W., T.R. Cary, D.J. Thomas, and C.D. Nickell 2006. Registration of 'LD00-3309' soybean. Crop Sci. 46:1384.

Diers, B.W., H.T. Skorupska, A.P. Rao-Arelli, and S.R. Cianzio. 1997. Genetic relationships among soybean plant introductions with resistance to soybean cyst nematode. Crop Sci. 37:1966–1972.

Dorrance, A. E., S.A. Berry, T.R. Anderson, and C. Meharg. 2008. Isolation, storage, pathotype characterization, and evaluation of resistance for *Phytophthora sojae* in soybean. Plant Health Prog. doi:10.1094/PHP-2008-0118-01-DG.

Faghihi, J., P.A. Donald, G. Noel, T.W. Welacky, and V.R. Ferris. 2010. Soybean resistance to field populations of *Heterodera glycines* in selected geographic areas. Plant Health Prog. doi:10.1094/PHP-2010-0426-01-RS.

Golden A.M., J.M. Epps, R.D. Riggs, L.A. Duclos, J.A. Fox, and R.L. Bernard. 1970. Terminology and identity of intraspecific forms of the soybean cyst nematode (*Heterodera glycines*). Plant Dis. Rep. 54:544-546.

Jackson, T., G. Smith, and T. Niblack. 2005. *Heterodera glycines* infectivity and egg viability following nonhost crops and during overwintering. J. Nematol. 37:259-264.

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

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

Kadam, S., T.D. Vuong, D. Qiu, C.G. Meinhardt, L. Song, R. Deshmukh, G. Patil, J. Wan, B. Valliyodan, A.M. Scaboo, J.G. Shannon, and H.T. Nguyen. 2016. Genomic-assisted phylogenetic analysis and marker development for next generation soybean cyst nematode resistance breeding. Plant Sci. 242:342-350.

Keim, P., T. Olson, and R. Shoemaker. 1988. A rapid protocol for isolating soybean DNA. Soybean Genet. Newslet. 12:150-152.

Kim, M., and B.W. Diers. 2013. Fine mapping of the SCN resistance QTL cqSCN-006 and cqSCN-007 from *Glycine soja* PI 468916. Crop Sci. 53:775-785.

Kim, M., D.L. Hyten, D. L., Bent, A. F., and Diers, B. W. 2010. Fine mapping of the SCN resistance locus from PI 88788. The Plant Genome. 2:81-89.

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

Koenning, S.R., and J.A Wrather. 2010. Suppression of soybean yield potential in the continental United States from plant diseases from 2006 to 2009. Plant Health Prog. doi:10.1094/PHP-2010-1122-01-RS.

Kopisch-Obuch, F., R. McBroom, and B. Diers. 2005. Association between soybean cyst nematode resistance loci and yield in soybean. Crop Sci. 45:956-965.

Li, X., Y. Han, W. Teng, S. Zhang, K. Yu, V. Poysa, T. Anderson, J. Ding, and W. Li. 2010. Pyramided QTL underlying tolerance to Phytophthora root rot in mega-environments from soybean cultivars ‘Conrad’and ‘Hefeng 25’. Theor. Appl. Genet. 121:651-658.

Matson, A.L., and L.F. Williams. 1965. Evidence of a fourth gene for resistance to the soybean cyst nematode. Crop Sci. 5:477.

Miller, D., S. Chen, P. Porter, G. Johnson, D. Wyse, S. Stetina, L. Klossner, and G. Nelson. 2006. Rotation crop evaluation for management of the soybean cyst nematode in Minnesota. Agron. J. 98:569-578.

Mudge, J., V. Concibido, R. Denny, N. Young, and J. Orf. 1996. Genetic mapping of a yield depression locus near a major gene for soybean cyst nematode resistance. Soybean Genet. Newslet. 23:175-178.

Niblack, T., P. Arelli, G. Noel, C. Opperman, J. Orf, D. Schmitt, G. Shannon, and G. Tylka. 2002. A revised classification scheme for genetically diverse populations of *Heterodera glycines*. J. Nematol. 34:279-288.

Niblack, T., A. Colgrove, K. Colgrove, and J. Bond. 2008. Shift in virulence of soybean cyst nematode is associated with use of resistance from PI 88788. Plant Health Prog. doi:10.1094/PHP-2008-0118-01-RS.

Niblack, T.L., and R.D. Riggs. 2015. Soybean cyst nematode. In: G.L. Hartman, J.C. Rupe, E.J. Sikora, L.L. Domier, J.A. Davis, and K.L. Steffy (eds), Compendium of soybean diseases (5th ed). APS Press, St. Paul, MN. p. 88-90.

Ortega, M.A., J.N. All, H.R. Boerma, and W.A. Parrott. 2016. Pyramids of QTLs enhance host–plant resistance and Bt-mediated resistance to leaf-chewing insects in soybean. Theor. Appl. Genet. 129:1-13.

Porter, P.M., S.Y. Chen, C.D. Reese, and L.D. Klossner. 2001. Population response of soybean cyst nematode to long term corn–soybean cropping sequences in Minnesota. Agron. J. 93:619-626.

Rao-Arelli, A., S.C. Anand, and J.A. Wrather. 1992. Soybean resistance to soybean cyst nematode race 3 is conditioned by an additional dominant gene. Crop Sci. 32:862-864.

SAS Institute. 2016. The SAS system for Windows. Release 9.4. SAS Institute, Cary, NC.

Schapaugh, W.T., P.A. Owen, K.M. Clark, and D.A. Sleper. 1998. Registration of 'Magellan' soybean. Crop Sci. 38:892-893.

Schmitt, D., and G. Shannon. 1992. Differentiating soybean responses to *Heterodera glycines* races. Crop Sci. 32:275-277.

Shi, A., P. Chen, D. Li, C. Zheng, B. Zhang, and A. Hou. 2009. Pyramiding multiple genes for resistance to soybean mosaic virus in soybean using molecular markers. Mol. Breed. 23:113-124.

Shi, Z., S. Liu, J. Noe, P. Arelli, K. Meksem, & Z. Li. 2015. SNP identification and marker assay development for high-throughput selection of soybean cyst nematode resistance. BMC Genomics. 16:314.

Tenuta, M., and A. Tenuta. 2015. Slowing the spread of soybean cyst nematode. Crops and Soils. 48:28-30. doi:10.2134/cs2015-48-5-7.

Tylka, G.L., and C.C. Marett. 2014. Distribution of the soybean cyst nematode, *Heterodera glycines*, in the United States and Canada: 1954 to 2014. Plant Health Prog. doi:10.1094/PHP-BR-14-0006.

Tylka, G., C. Marett, and M. Serrano. 2015. Field experiments show effects of Clariva™ seed treatment in 2014. Integrated Crop Management News, Iowa State University.

University of Illinois. 2016. Soybean variety test results in Illinois: 2016 Report. Variety Testing Department of Crop Sci. http://vt.cropsci.illinois.edu/soybean.html (accessed Dec 7 2016).

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

Williamson, V.M., and R.S. Hussey. 1996. Nematode pathogenesis and resistance in plants. The Plant Cell. 8:1735-1745.

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

Wang, D., J. Shi, S. Carlson, P. Cregan, R. Ward, and B. Diers. 2003. A low-cost, high-throughput polyacrylamide gel electrophoresis system for genotyping with microsatellite DNA markers. Crop Sci. 43:1828-1832.

Wu, X., S. Blake, D.A. Sleper, J.G. Shannon, P. Cregan, and H.T. Nguyen. 2009. QTL, additive and epistatic effects for SCN resistance in PI 437654. Theor. Appl. Genet. 118:1093-1105.

Young, L.D. 1999. Soybeans resistant to *Heterodera glycines* populations attacking Hartwig cultivar. J. Nematol. 31:583.

Yu, N., and B.W. Diers. 2016. Fine mapping of the SCN resistance QTL cqSCN-006 and cqSCN-007 from *Glycine soja* PI 468916. Euphytica. In press.

APPENDIX A: SOYBEAN CYST NEMATODE SUPPLEMENTAL MATERIALS

Tables and Figures

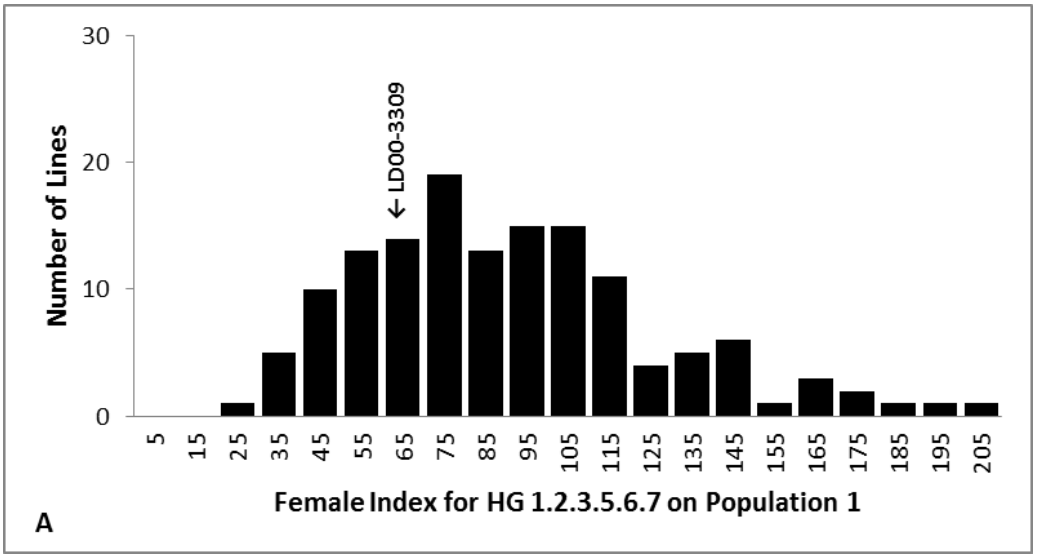
Table A.1. Genotypes of lines in Population 1 based on markers linked *rhg1* and the resistance quantitative trait loci (QTL) on chr 10 (CHR10). R indicates homozygous for the resistance allele of the QTL while S indicates homozygous susceptible at the QTL.

QTL†		Number of lines
rhg1	CHR10	
R	R	39
R	S	37
S	R	37
R	R	32

Table A.2. Genotypes of lines in Population 2 based on markers linked *rhg1*, cqSCN-007, cqSCN-007, the resistance quantitative trait loci (QTL) on chr 10 (CHR10). R indicates homozygous for the resistance allele of the QTL while S indicates homozygous susceptible at the QTL.

QTL†				Number of Lines
rhg1	cqSCN-006	cqSCN-007	CHR10	
S	S	S	S	4
R	S	S	S	12
S	R	S	S	7
S	S	R	S	8
S	S	S	R	9
R	R	S	S	4
S	R	R	S	9
S	S	R	R	4
R	S	S	R	9
R	S	R	S	8
S	R	S	R	3
R	R	R	S	4
S	R	R	R	4
R	S	R	R	8
R	R	S	R	10
R	R	R	R	4

Figure A.1. Distribution of female index (FI) for soybean cyst nematode (SCN) isolates HG 1.2.3.5.6.7 and TN21 in Population 1, which segregates for resistance quantitative trait loci (QTL) from PI88788 and PI567516C (A-B), and Population 2, which segregates for resistance QTL from PI88788, PI567516C, and PI468916 (C-D). Distribution of FI for SCN isolates HG 2.5.7 and TN23 on the Population 2 Subset, which contains selected lines from Population 2 (E-F). The average female index values of the recurrent parent LD00-3309 are denoted by an arrow.



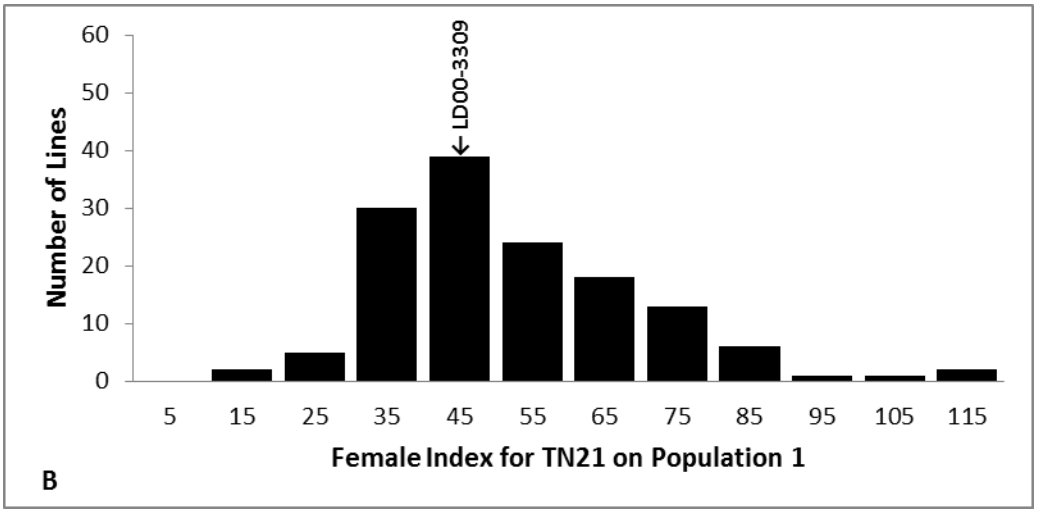
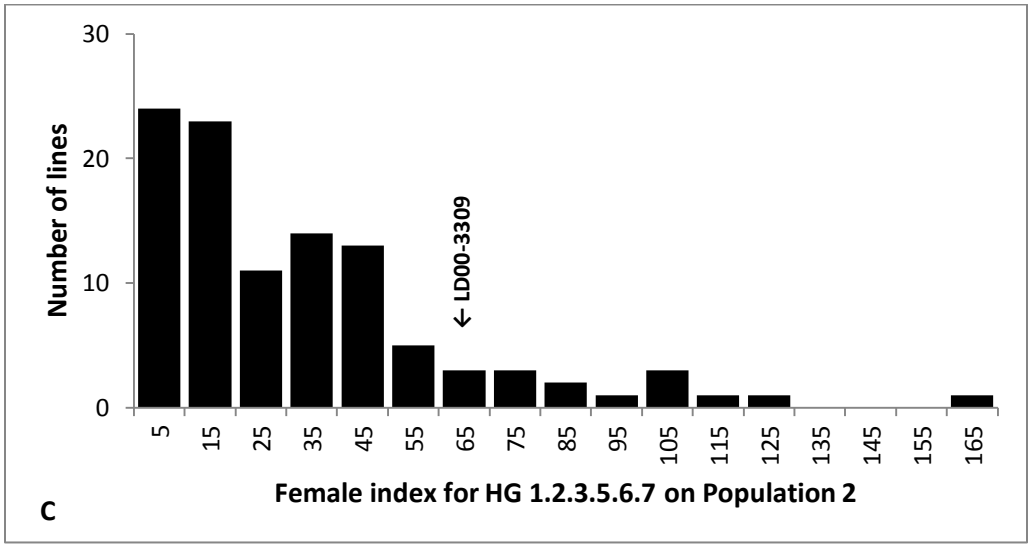


Figure. A.1. (cont.)



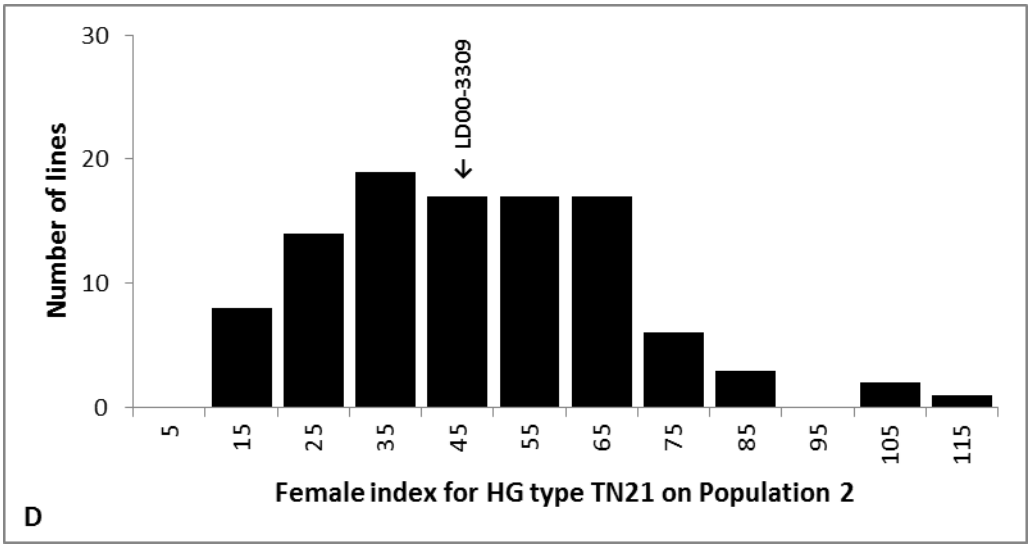


Figure. A.1. (cont.)