Genome-Wide Analyses for Partial Resistance to *Phytophthora sojae* Kaufmann and Gerdemann in Soybean (*Glycine max* L. Merr.) Populations from North America and the Republic of Korea

THESIS

Presented in Partial Fulfillment of the Requirements for the Degree Master of Science in the Graduate School of The Ohio State University

By

Rhiannon Noelle Schneider

Graduate Program in Horticulture and Crop Science

The Ohio State University

2015

Master's Examination Committee:

Dr. Leah K. McHale, Advisor
Dr. Anne E. Dorrance, co-Advisor
Dr. Clay H. Sneller
Abstract

Phytophthora root and stem rot of soybean (Glycine max) is caused by the oomycete pathogen Phytophthora sojae. This disease can be controlled by genetic resistance, but can cause devastating yield losses in fields planted with susceptible soybean cultivars and results in losses of around $300 million annually in the US. Partial resistance is considered to be more durable against P. sojae than race-specific resistance conferred by Rps genes and is theoretically effective against all races of this pathogen. Evaluation of a historical set of public cultivars representing 80 years of soybean breeding indicated that there have been genetic gains for partial resistance; however, these gains may have begun to plateau in the 1970s to early 1980s. Cultivars developed in Ohio generally have high levels of partial resistance to P. sojae; however, there is little known about the genetic regions associated with the partial resistance. Further improvement of increasing partial resistance could be achieved through the introgression of known quantitative trait loci (QTL) from plant introductions from the Republic of Korea (South Korea), which contain high levels of partial resistance. From an analysis of 1,398 plant introductions with a wide range of phenotypic expression of resistance, sixteen single nucleotide polymorphisms (SNPs) were associated with partial resistance to P. sojae. These SNPs were located in three genomic regions, or QTL, on chromosomes 3, 13, and 19. The QTL on chromosome 19 represented a novel locus, whereas the QTL on chromosomes 3 and 13 were
coincident with previously identified QTL for partial resistance and/or Rps genes. In contrast, a genome-wide association study carried out in Ohio breeding lines was unable to detect any significant marker-trait associations, limiting the ability to use marker assisted selection to improve partial resistance in this population. However, genomic selection (GS) was shown to be a promising means of selection, with efficiencies relative to phenotypic selection of 0.5 to 1. Importantly, GS can be implemented through use of multi-trait indices which include yield. As exotic germplasm with high levels of partial resistance are identified, GS may be a valuable tool for utilizing exotic sources of partial resistance to P. sojae while maintaining or improving yield.
Dedication

To my mother and father, who taught me to never give up on my dreams.
Acknowledgements

First, I would like to thank Dr. Leah K. McHale for enlightening me with her wisdom about plant breeding and genetics throughout my undergraduate and graduate years. I appreciate her guidance and support as a mentor to me with my thesis project, in addition to allowing me the opportunity for an internship with Pioneer Hi-Bred towards the end of my graduate study. She has always been an inspiring role model to me, and I could not be grateful enough to have such an amazing advisor because I would not be where I am today if it was not for her.

I would like to thank my co-advisor, Dr. Anne E. Dorrance for teaching me all there is to know about plant pathology, especially dealing with soybeans! Her guidance and helpful suggestions while I was conducting disease assays in Wooster was a tremendous learning experience for me.

In addition, I would like to thank Dr. Clay H. Sneller for his knowledgeable input about association mapping and genomic selection for my project and for being a member on my committee.

I have much gratitude and thanks to the Dorrance lab, who assisted me with the large, time-consuming disease assays; including Jaqueline Huzar Novakowiski, Charlotte Smith, and Deloris Veney and former lab members Andika Gunadi and Sungwoo Lee. Special thanks to the undergraduates: Lisa Sutton, Sarah Lewis, Allen Honerlaw, and Tom Fitz Gibbon. I also owe a huge thank you to the McHale lab for assisting me with my project, especially Amanda Gutek, Angie Parker, Christine Dubler, Scott McIntyre and former lab members Keith Freewalt, Mao Huang, Bradley Snyder, and Elizabeth Baskin.

I would like to acknowledge the Ohio Soybean Council for awarding me with an OSC scholarship during my undergraduate and graduate studies.
Finally, I would like to thank all of my friends, family, and my significant other, Branden Overstreet for their love and support. I am truly blessed and privileged to have these special people in my life.
Vita

August, 1987.................................Born: Middleburg Heights, OH, U.S.A.

2012..............................................B.S in Agriculture, The Ohio State University,
                               Columbus, OH

2012 to present..............................Graduate Research Associate, Department of
                               Horticulture and Crop Science,
                               The Ohio State University

2014 to present..............................Research Associate, DuPont Pioneer Hi-bred
                               Napoleon, Ohio

Publications

Peer-reviewed articles:

Chakrabarti M, Zhang N, Sauvage C, Muños S, Blanca J, Cañizares J, Diez MJ, Schneider R,
                               regulates a domestication trait in cultivated tomato. Proceedings of the National

Poster presentations:

Schneider R, Dorrance A, McHale L (2014) Genome wide association study for partial resistance
                               to Phytophthora sojae in soybean plant introductions from South Korea. Horticulture
                               and Crop Science Graduate Retreat, Wooster, OH

                               in a diverse collection of soybean cultivars, breeding lines, and plant introductions. ASA-
                               CSSA-SSSA Annual meetings, Tampa, FL. P115

Schneider R, McHale L (2012) Marker Development for marker assisted breeding of resistance to
                               Phytophthora sojae in soybean. The Ohio State University. Denman Undergraduate
                               Research Forum, Columbus, OH.

Fields of Study

Major Field: Horticulture and Crop Science
Table of Contents

Abstract............................................................................................................................... ii

Dedication............................................................................................................................. iv

Acknowledgements.............................................................................................................. v

Vita........................................................................................................................................ vii

List of Tables .......................................................................................................................... ix

List of Figures ......................................................................................................................... x

Chapter 1: Literature Review ............................................................................................... 1

Chapter 2: Genetic Gains for Partial Resistance to *Phytophthora sojae* in Maturity Group II and III Public Soybean Cultivars.............................................................................................................................. 13

Chapter 3: Genome-Wide Association Mapping and Genomic Selection for Partial Resistance to *Phytophthora sojae* in a Soybean Breeding Population...................................................................................................................... 26

Chapter 4: Genome-Wide Association Mapping of Partial Resistance to *Phytophthora sojae* in Soybean Plant Introductions from the Republic of Korea................................................. 52

Bibliography .......................................................................................................................... 76
List of Tables

Table 1.1. Summary of QTL identified in crosses with North American lines .................................. 11

Table 1.2. Summary of QTL identified in Plant Introductions from the Republic of Korean .......... 12

Table 3.1. Summary of mean lesion lengths from tray tests and virulence profiles from hypocotyl inoculations utilizing 23 different isolates of P. sojae ................................................................. 45

Table 3.2. Heritability and genetic variance of each trait ................................................................. 46

Table 3.3. Mean prediction accuracies of genomic selection ............................................................ 47

Table 3.4. Prediction accuracies and phenotypic correlations between isolates and relative
efficiency of genomic selection ...................................................................................................... 48

Table 4.1. Variance components and heritability ($H^2$) of each trait .............................................. 68

Table 4.2. Significant marker-trait associations with identified genomic regions and chromosome location ................................................................................................................................. 69

Table 4.3. South Korean plant introductions with the highest levels of root score resistance .... 70
List of Figures

Figure 2.1. Histogram, scatterplot, and correlation coefficients for inoculated traits in the North American soybean population .................................................................................................................................................. 23

Figure 2.2. Simple linear regression and piecewise linear regression for inoculated traits .......... 24

Figure 2.3. Simple linear regressions for non-inoculated traits .................................................................................................................................................................................. 25

Figure 3.1. Histogram, scatterplot, and Pearson’s correlation for inoculated traits in the Ohio soybean population ............................................................................................................................................................................. 49

Figure 3.2. Overlapping histograms with the raw values of the four inoculated traits for isolates OH121086.3 and C2S1 ................................................................................................................................................................................. 50

Figure 3.3. Manhattan plot from genome-wide association study for the OSU population ....... 51

Figure 4.1. Histogram, scatterplot, and Pearson’s correlation for inoculated traits in soybean plant introductions from the Republic of Korea .................................................................................................................................................................. 71

Figure 4.2. PCA analysis plot for plant introductions from the Republic of Korea .................. 72

Figure 4.3. Manhattan plot from genome-wide association study for plant introductions from the Republic of Korea for inoculated traits ............................................................................................................................................................................. 73

Figure 4.4. Manhattan plot from genome-wide association study for plant introductions from the Republic of Korea for non-inoculated traits ..................................................................................................................................................................... 74

Figure 4.5. Manhattan plot of significant marker-trait associations and previously indentified Rps gens and QTL for partial resistance to Phytophthora sojae ................................................................................................................................................. 75
Chapter 1

LITERATURE REVIEW

Introduction

Phytophthora stem and root rot (PSR) is a destructive disease in soybean, caused by *Phytophthora sojae*, a soil-borne oomycete pathogen that thrives in poorly drained, wet soils (Schmitthenner, 1985). PSR has been attributed to soybean yield losses of around $300 million annually in the United States (Wrather and Koenning 2006). This destructive pathogen is found in soybean producing regions throughout the US and worldwide (Schmitthenner, 1985; Anderson and Buzzell 1992; Yanchun and Chongyao 1993; Jee et al. 1998; Dorrance and Grunwald 2009; Grau et al. 2004; Lee et al. 2013a). The asexual zoospores can infect soybean roots throughout the growing season (Schmitthenner, 1985). Although PSR was first reported in 1948 in Indiana, *P. sojae* was not identified until 1958 by Kaufmann and Gerdemann as the causal pathogen of PSR (Schmitthenner, 1985). Application of chemical treatments is one option for controlling this pathogen; however, this method can have negative impacts on yield when *P. sojae* is absent and is not completely effective when implemented as the only management practice (Schmitthenner, 1985). The use of resistant cultivars is considered a more efficient method for managing this pathogen. Resistance to *P. sojae* can be inherited in either a qualitative or quantitative manner, both of which can be combined in a single cultivar. Additionally, the presence of both types of resistance does not have a negative impact on yield when the pathogen is absent (Tooley and Grau, 1984a; St. Martin et al., 1994; Dorrance et al., 2003a).
Qualitative resistance is controlled by single, dominant genes, known as Resistance to Phytophthora sojae (Rps) genes. This manner of resistance provides complete protection by inducing programmed cell death through the plant’s defense signaling pathway, which prevents the infection from spreading (Zhang et al., 2011). This race-specific response is triggered by effector proteins coded by Avirulence (Avr) genes, which are secreted by P. sojae zoospores into the soybean root. In plant systems that have R – Avr gene interactions, effector proteins can be directly or indirectly recognized by an R gene acting as a receptor and signal transducer of defense. A “gene-for-gene” interaction occurs when an incompatible reaction between the Avr genes of the pathogen and the Rps genes present in the soybean (Shan et al., 2004). This interaction places strong selection pressure on the P. sojae population, such that a mutation or silencing of an Avr gene allows the pathogen to evade reception by the plant’s cognate Rps protein and rapidly increase in frequency within the population. Consequently the P. sojae population evolves into a new, virulent race, where the Rps genes are no longer effective against the pathogen. This may account for P. sojae populations exhibiting evidence of rapid evolution and high diversity (Dong et al., 2009). Currently, there are 55 physiological races of P. sojae identified in the U.S., which excludes P. sojae pathotypes with a compatible interaction to all deployed Rps genes (Schmitthenner et al., 1994; Yang et al., 1996; Abney et al. 1997; Leitz et al., 2000; Kaitany et al., 2001; Dorrance et al., 2003b). The effectiveness of a single Rps gene has been estimated to last from eight to fifteen years (Schmitthenner, 1985). This limited lifespan, as well as the highly diverse pathogen population, highlights the need for non-race-specific resistance.

Quantitative resistance, or partial resistance, is theoretically non-race specific and therefore expected to provide a broad spectrum of resistance against all races of P. sojae
This type of resistance is considered more stable and durable against this aggressive pathogen (Tooley and Grau, 1984a,b; Schmitthenner, 1985). The stability of partial resistance can be attributed to its quantitative inheritance resulting from the action of multiple genes each contributing a small measure of resistance and to an overall reduction in disease development (Van der Plank, 1968; Umaerus, 1970; Poland et al., 2009; St. Clair et al., 2010). Infected phenotypes with high levels of partial resistance often display reduced disease symptoms, such as a reduction in the number of successful infections, reduction in overall root rot and stem rot and the amount of pathogen colonization and reproduction, as well as an increase in the latent period of infection (Young, 1996; Kou and Wang, 2010; Mideros et al., 2007). The use of cultivars with high levels of partial resistance can slow the evolution of the pathogen due to the reduced selection pressure conferred by quantitative trait loci (QTL) as compared to Rps genes. Similar to Rps genes, the high levels of partial resistance to P. sojae can effectively manage this disease.

It has been posited that the Republic of Korea (South Korea) is the origin of the P. sojae-soybean pathosystem and that the S. Korean soybean germplasm has co-evolved with P. sojae due to the high levels of partial resistance and diversity of resistance phenotypes in collections from this geographic region (Dorrance and Schmitthenner, 2000; Gordon et al., 2006). Thus, high genetic diversity for resistance to P. sojae is also expected to exist within soybean Plant Introductions (PIs) from S. Korea. These S. Korean PIs could be used as sources of partial resistance and key loci introgressed into North American breeding lines. In order to do so in an informed manner, the resistance alleles which are currently present in the North American germplasm and contributing to the partial resistance to P. sojae should first be characterized. Thus far, the QTL studies in N. American cultivars have largely been limited to resistance
contributed from the cultivar ‘Conrad’ (Burnham et al., 2003; Weng et al., 2007; Han et al., 2008; Li et al., 2010; Wang et al., 2010; Wang et al., 2012). Genome-wide association (GWA) mapping may provide information on common alleles conferring resistance to P. sojae in natural populations of S. Korean or N. American germplasm.

While identifying and comparing QTL in S. Korean and N. American populations has utility in improving N. American breeding lines through marker-assisted introgression of resistance alleles from the S. Korean germplasm, it is also crucial to develop a breeding strategy suitable for the concurrent selection of numerous QTL. Genomic selection (GS) utilizes high-density markers that cover the entire genome that are not from a predetermined or defined set of significant markers (Heffner et al., 2009). The GS model is generated from a “training population” that has both phenotypic and genotypic data in order to determine genomic estimated breeding values (GEBV) (Meuwissen et al., 2001, Habier et al., 2007; Heffner et al., 2009). Progeny are then selected by their genotypes based on their GEBVs, eliminating phenotypic selection. This gives GS the advantage over phenotypic selection by eliminating the need for several breeding cycles to take place, which necessarily includes several generations of inbreeding and seed increase. GS can also maintain genetic diversity within the breeding program, and can increase genetic gains beyond the capacity of phenotypic selection and QTL approaches (Heslot et al., 2012). GS has been shown to be more effective than marker assisted selection (MAS) due to the utilization of all the available markers, regardless of the significance of their effect, versus a subset of markers with a significant effect (Massman et al. 2013). With its shorter timeframe in comparison to phenotypic selection and increased marker coverage relative to MAS, GS may be a viable method of increasing genetic gains for P. sojae resistance.
Partial resistance to \textit{P. sojae}

A number of studies have identified QTL in N. American (Table 1) and S. Korean (Table 2) germplasm. QTL studies of N. American lines with high levels of partial resistance have been primarily limited to the cultivar Conrad. Two QTL contributing 30 to 50\% of the phenotypic variation were identified in Conrad using three Recombinant Inbred Line (RIL) populations (Conrad (resistant, R) x Sloan (susceptible, S), Conrad x Harsoy (S), and Conrad x Williams (S)) in a tray test. The QTL were positioned on chromosome 13 (molecular linkage group (MLG) F) and chromosome 2 (MLG D1b+W) (Burnham et al., 2003). Five QTL were identified on chromosomes 12 (MLG H), 13, 14 (MLG B2), 17 (MLG D2), and 19 (MLG L) using a tray test (Wang et al., 2010). Another QTL was identified on chromosome 16 in a F\textsubscript{6} RIL population (Conrad x OX760-6-1 (S)) (Weng et al., 2007). In an F\textsubscript{7} RIL population also derived from Conrad x OX760-6-1 (S), Han et al. (2008) detected three QTL, two were located on chromosome 13 (MLG F) and the third was located on chromosome 2 (MLG D1b). Additional QTL were identified in a population derived from a cross between Conrad and a Chinese cultivar, ‘Hefeng 25’ (R). Through multiple field trials, eight QTL were attributed to ‘Conrad’ (located on chromosomes 2, 6 (MLG C2), 8 (MLG A2), 11 (MLG B1), and 13) and five QTL to ‘Hefeng 25’ (Li et al. 2010). Three QTL were identified in an interspecific population derived from a cross between Virginia soybean cultivar (R), ‘V71-370’ and a \textit{Glycine sojae} (wild soybean species), ‘PI 407162’ (S), which were phenotyped via a tray test. One major and minor QTL were associated with the PI 407162 allele on chromosomes 16 (MLG J) and 18 (MLG G), and one minor QTL was associated with V71-370 on chromosome 20 (MLG I) (Tucker et al., 2010).

Although Plant Introductions (PI) from regions in South Korea and China have been previously evaluated for their level of partial resistance (Dorrance and Schmitthenner 2000),
studies identifying QTL have been limited to seven PIs. In a series of recent studies, six RIL soybean populations were derived from crosses of ‘OX20-8’, a breeding line with Rps1a and no partial resistance, and PIs originating from China, Korea, and Japan for identification of QTL to P. sojae. The study found a total of 50 QTL mapped to approximately 25 locations across all chromosomes. Except for chromosome 11, chromosomes 1 (MLG D1a), 2, 3 (MLG N), 13, 14 (MLG B2), 16, 17 (MLG D2), and 18 (MLG G) were recognized as regions with high levels of partial resistance to P. sojae (Lee et al. 2013a; Lee et al. 2013b; Lee et al. 2014). In the population derived from a cross between OX20-8 and PI 407861A, which originated from S. Korea, nine QTL were identified on chromosomes 3, 4 (MLG C1), 8, 10 (MLG O), 13, 15 (MLG E), and 18, all located near known R-gene rich regions and previously stated QTL to soil-borne pathogens. Of the nine, two were first identified as novel QTL to P. sojae in this study and were located on chromosomes 3 and 8 (Lee et al., 2013b). In the population derived from a cross between OX20-8 and PI 398841, which also originated from S. Korea, a total of ten QTL were identified. Seven of these QTL were first reported as novel and were found near loci of previously reported Rps, isoflavone and oil content genes, as well as loci contributing resistance to other soil-borne fungal pathogens. The remaining three QTL were co-localized with known Rps genes on chromosomes 3, 13, and 18; however, following a hypocotyl inoculation assay, PI 398841 did not give a typical R-gene response (Lee et al., 2013a). Nguyen et al. (2012) identified two QTL associated with both flooding tolerance and with partial resistance to P. sojae, located on chromosome 11 and 13. These studies suggest that PIs from S. Korea can be used as sources for improvement of partial resistance to P. sojae in N. American cultivars.
Genome-wide association mapping

GWA mapping can be used to complement bi-parental mapping and, due to increased precision, can resolve questions related to lineage and pleiotropic traits. When breeding populations are used, GWA mapping can identify genetic markers suitable for immediate crop improvement (Massman et al., 2011). GWA mapping has been recently used as a tool towards marker assisted selection (Varshney et al., 2012). One of the first GWA plant studies was completed on the plant model, Arabidopsis thaliana, in order to test GWA mapping and its efficiency (Aranzana et al., 2005). The study’s objective was to look at known QTL affecting flower time and disease resistance in 95 accessions. Although there were a high number of false positives resulting from population structure, the study was able to identify known genes contributing to flowering time and provided evidence for the potential application of GWA mapping in other plant species (Aranzana et al., 2005). A later study demonstrated that controlling for population structure using a mixed model approach, which included marker effect, a kinship matrix (K) and population structure (Q) (Yu et al. 2006), was the most sufficient compared to the eight other models used; naïve, Q, K, K* (using an alternative kinship matrix), Q+K*, principal component analysis (PCA), PCA+K, and P+K*. Simulations have also shown that population sizes greater than 384 individuals are required to consistently detect QTL (Wang et al. 2012).

GWA studies have since been conducted on a number of crop species including barley (Hordeum vulgare) (Varshney et al. 2012), maize (Zea mays) (Zhang et al. 2013), sugar beets (Beta vulgaris) (Wurschum et al. 2011), soybean (Mamidi et al. 2011), and apple (Malus domestica) (Li et al. 2011) for drought tolerance, flooding tolerance, sugar yield traits, iron deficiency chlorosis, and fruit development. These later studies have been able to confirm...
previously identified QTL. For example, Belo et al. (2008) detected a single major QTL influencing oleic oil content in a population of 553 maize inbred lines, which confirmed a locus identified in a previous QTL mapping study (Alrefai et al., 1995). GWA mapping studies have also identified novel QTL; e.g. GWA mapping for kernel and malting quality traits in barley identified 15 novel QTL out of 140 significant marker-trait associations for 11 out of 19 traits (Matthies et al., 2014). In addition, GWA mapping studies have quickly identified candidate genes for further functional gene analyses, such as Xue et al. (2013), identifying five candidate genes conferring drought tolerance in maize by identifying single nucleotide polymorphisms (SNPs).

Genomic Selection

Genomic (genome-wide) selection (GS) utilizes the advancement of higher throughput marker technologies to generate genomic estimated breeding values (GEBV) such that each cycle of selection can be shortened relative to a cycle of phenotypic selection (PS). Meuwissen et al. (2001) first proposed the utilization of high-density marker coverage in GS for increased efficiency in the selection of large breeding populations. Genomic selection models have been previously evaluated in several simulation and empirical studies. Shortening each breeding cycle of selection is the primary advantage of genomic selection. This was especially apparent in a simulation study on oil palm (Wong and Bernardo, 2008), where one cycle of PS in oil palm requires nineteen years. Phenotypic selection, marker assisted recurrent selection (MARS), and genomic selection were compared for selection gain per unit time and cost by simulation using small population sizes of oil palm. They found the selection response (based on the genetic standard deviation in cycle 0) for PS and GS to be greater than the response to MARS, however, responses to GS were 4-25% larger than the coinciding response to PS with a population size of
50 or 70. In addition, the cost per unit gain was 26-57% lower with GS than with PS when markers cost $1.50 per data point, as well as GS being 2-3 years less than PS in time per unit gain (Wong and Bernardo 2008). An empirical study (Heffner et al., 2010) also evaluated genetic gains for a wheat breeding program by comparing PS, MAS, and GS prediction accuracies for 13 agronomic traits in 374 winter wheat advanced cycle breeding lines. Using a cross validated approach; the average prediction accuracies found in GS were 28% greater than with MAS and were 95% of PS. For net merit prediction accuracy (trait predictions plus a weighting index) the average accuracy across six selection indices for GS was 14% greater than PS. These studies provide evidence for the idea that GS can greatly accelerate a breeding program, but the statistical methods selected to use for GS can have varying results.

In a broad study of GS models, eight wheat, barley, Arabidopsis, and maize datasets were collected for studying the predictive ability of currently available GS models along with several machine learning (ML) methods (Heslot et al., 2012). The GS models and ML methods were evaluated by genomic estimated breeding values (GEBV) and marker effects for each model. Many models resulted in a similar level of accuracy, conversely, levels of over fitting, computation time, and distribution of marker effect estimates widely varied. However, the Bayesian Lasso, the Ridge Regression –Best Linear Unbiased Predictor, and the weighted Bayesian shrinkage regression models were recommended for use in a breeding program (Heslot et al., 2012).

Hypotheses and Objectives

It is hypothesized that 1) there have been historic genetic gains for partial resistance to P. sojae, 2) utilization of marker-assisted selection or genomic selection are feasible methods for
The objectives included:

1) Estimate genetic gains for resistance to *P. sojae* over the past 80 years of soybean breeding.

2) Conduct genome-wide association studies in two populations, a population of soybean cultivars and breeding lines from Ohio and a population of plant introductions which originated from S. Korea.

3) Compare three different GS models derived from a compressed mixed linear model (CMLM), mixed linear model (MLM), and enriched CMLM (ECMLM) by cross-validation.

4) Evaluate the efficiency of GS relative to phenotypic selection for partial resistance to *P. sojae*. 

improvement of partial resistance to *P. sojae*, and 3) QTL associated with partial resistance to *P. sojae* can be identified in natural populations.

The aim of this project was to detect QTL in soybean conferring partial resistance towards *P. sojae* in N. American and S. Korean populations while identifying common alleles between the two populations. In addition to determining the number and location of QTL in the soybean genome, a viable method for improving partial resistance towards *P. sojae* in OSU-OARDC breeding program.

The objectives included:

1) Estimate genetic gains for resistance to *P. sojae* over the past 80 years of soybean breeding.

2) Conduct genome-wide association studies in two populations, a population of soybean cultivars and breeding lines from Ohio and a population of plant introductions which originated from S. Korea.

4) Compare three different GS models derived from a compressed mixed linear model (CMLM), mixed linear model (MLM), and enriched CMLM (ECMLM) by cross-validation.

5) Evaluate the efficiency of GS relative to phenotypic selection for partial resistance to *P. sojae*. 


### Table 1.1. Summary of QTL identified in crosses with North American lines.

<table>
<thead>
<tr>
<th>Studies with North American lines</th>
<th>Number of QTL identified</th>
<th>QTL location chromosome</th>
<th>Population/Parents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burnham et al. 2003</td>
<td>2</td>
<td>2 &amp; 13</td>
<td>Conrad (R) x Sloan (S'), Williams(S), &amp; Harosoy (S)</td>
</tr>
<tr>
<td>Han et al. 2008</td>
<td>2</td>
<td>2 &amp; 13</td>
<td>Conrad x OX760-6-1 (S)</td>
</tr>
<tr>
<td>Weng et al. 2007</td>
<td>1</td>
<td>16</td>
<td>Conrad x OX760-6-1</td>
</tr>
<tr>
<td>Wang et al. 2010;</td>
<td>5</td>
<td>12, 13, 14, 17, &amp; 19</td>
<td>Conrad x Sloan</td>
</tr>
<tr>
<td>Wang H. et al. 2012</td>
<td>2</td>
<td>18 &amp; 19</td>
<td>Conrad x Sloan</td>
</tr>
<tr>
<td>Li et al. 2010</td>
<td>13</td>
<td>2, 6, 8, 11 &amp; 13</td>
<td>Conrad x Hefeng25 (R)</td>
</tr>
<tr>
<td>Wu et al. 2011</td>
<td>3</td>
<td>6, 10, &amp; 15</td>
<td>Su88-M21(R)x Xinyixiaohedou (S)</td>
</tr>
</tbody>
</table>

† R: high levels of partial resistance

‡ S: susceptible
<table>
<thead>
<tr>
<th>Study with Plant Introductions</th>
<th>Number of QTL identified</th>
<th>QTL location chromosomes</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tucker et al., 2010</td>
<td>3</td>
<td>16, 18, &amp; 20</td>
<td>V71–370 (R†) x PI 407162 (S‡)</td>
</tr>
<tr>
<td>Nguyen et al., 2012</td>
<td>2</td>
<td>13 &amp; 17</td>
<td>S99-2281 (S) x PI 408105A (R)</td>
</tr>
<tr>
<td>Lee et al., 2013b</td>
<td>9</td>
<td>3, 4, 8, 10, 13, 15 &amp; 18</td>
<td>OX20-8(S) x PI 407861A(R)</td>
</tr>
<tr>
<td>Lee et al., 2013a</td>
<td>10</td>
<td>1, 2, 3, 4, 7, 15, 18 &amp; 20</td>
<td>OX20-8(S) x PI 398841(R)</td>
</tr>
<tr>
<td>Lee et al., 2014</td>
<td>4</td>
<td>4, 9, 12, 16 &amp; 18</td>
<td>OX20-8 x PI 398841, PI 407861A, PI 427106 (R), PI 427105B (R), PI 398297 (R), PI 417178 (R)</td>
</tr>
</tbody>
</table>

**Table 1.2.** Summary of QTL identified in Plant Introductions from the Republic of Korean.

† R: high levels of partial resistance
‡ S: susceptible
Chapter 2

Genetic Gains for Partial Resistance to *Phytophthora sojae* in Maturity Group II and III Public Soybean Cultivars.

**Abstract**

First discovered in 1948, *Phytophthora* root and stem rot, caused by the oomycete *Phytophthora sojae*, is a destructive disease of soybean (*Glycine max* (L.) Merr). The primary measure of control has been through the utilization of soybean varieties with genetic resistance. Qualitative and quantitative are two distinct forms of genetic resistance, where qualitative resistance is characterized by a single gene versus quantitative resistance being controlled by more than one gene. Soybean cultivars with race-specific resistance against *P. sojae* were discovered around the same time the pathogen was first identified, marking the initial efforts of breeders targeting resistance to *P. sojae (Rps)* genes. While breeding efforts have purposefully targeted a range of *Rps* genes, breeding for quantitative or partial resistance against *P. sojae* has been less documented. This study examined partial resistance to *P. sojae* in a historic set of 38 and 57 soybean public cultivars in maturity groups (MG) II and III, respectively, released from 1923 to 2008 to determine trends in partial resistance to *P. sojae* over time. A regression of traits for partial resistance against year of cultivar released indicates that modern cultivars have higher levels of partial resistance to *P. sojae* than older cultivars.
Introduction

Phytophthora root and stem rot was first identified as a disease of soybean [Glycine max (L.) Merr] in Indiana 1948, but was not associated with the oomycete pathogen Phytophthora sojae until 1954, the same year it was found in North Carolina (Schmitthenner, 1985). Kauffman and Gerdemann (1958) were the first to classify this pathogen as Phytophthora sojae. Since then, the pathogen has gone through a series of name changes (Kuan and Erwin, 1980), but is widely known as P. sojae.

Multiple races of P. sojae have been identified in the Northern US beginning with race 3 (1972), race 4 (1974), race 5 and 6 (1976), and race 7, 8, and 9 in 1978 (Schmitthenner, 1985). Each race is defined on the basis of its differential reaction to germplasm containing resistance to P. sojae (Rps) genes. Only 55 physiological races of P. sojae have been named in the US, while many P. sojae pathotypes with a compatible interaction to all deployed Rps genes have been found (Schmitthenner et al., 1994; Yang et al., 1996; Abney et al., 1997; Leitz et al., 2000; Kaitany et al., 2001; Dorrance et al., 2003; Dorrance et al., data unpublished). Although the origin of P. sojae is unknown, it has been hypothesized that P. sojae co-evolved with soybean in South Korea (Gordon et al., 2007).

P. sojae became a major pathogen of soybean in a relatively short period of time (Schmitthenner, 1985). Phytophthora root and stem rot causes annual yield losses in excess of $300 million in the United States (Wrather and Koenning, 2006). Management for this pathogen is primarily through quantitative (partial) or qualitative genetic resistance. Qualitative resistance is race-specific and is conferred by single, dominant Rps genes. Resistance to P. sojae in soybean was first discovered in 1954 in two northern US cultivars of maturity group (MG) I possessing Rps1 (Schmitthenner, 1985). Since its discovery, improvement of P. sojae resistance in N.
American soybean cultivars has been an important focus of breeding programs. Currently, 20 Rps genes have been identified (Lin et al., 2013) and have been or are being introgressed into soybean cultivars in an effort to control this constantly evolving pathogen (e.g. Orf et al., 1987; Kilen et al., 1995; Nickell et al., 1996; St. Martin et al., 2008; McHale et al., 2013).

In contrast, partial resistance does not provide the complete immunity conferred by Rps genes; however, it is theoretically non-race specific. While partial resistance has been considered a more durable resistance due to its effectiveness against a wide range of P. sojae races (Tooley and Grau, 1984; Schmitthenner, 1985), its quantitative nature makes it a less desirable target for breeding as compared to Rps genes. In addition, partial resistance can be masked by race-specific resistance. These factors combine to make it unclear whether genetic gains have been made for partial resistance to P. sojae. The objective of this study was to determine if there have been significant changes in partial resistance over the 80 years of soybean breeding in the U.S. Ninety-five MG II and III plant introductions (PIs) and public cultivars released between 1923 and 2008, which were widely grown by soybean producers after their release, were selected for inclusion in this study (Rincker et al., 2014).

Materials and Methods

Plant Materials

The experimental group consisted of 95 PIs and public cultivars (MG II or III) which were harvested from the OARDC-OSU research station in Wooster, OH in 2012. Eight cultivar checks with known levels of partial resistance were included as checks in the phenotypic disease assay. All of the seed was surface sterilized following a chlorine gas protocol adapted from Olhoft et al. (2006).
**P. sojae Isolates**

*P. sojae* isolates OH12108_6.3 (OH121) (vir 1a, 1b, 1c, 1d, 1k, 2, 3a, 3c, 4, 5, 6, 7, and 8) and C2S1 (vir 1a, 1b, 1d, 1k, 2, 3a, 3b, 3c, 4, 5, 6, 7, and 8) were selected from pathogenicity screens of 27 isolates for virulence and aggressiveness by hypocotyl and tray tests, respectively (Olah and Schmitthenner, 1985; Dorrance et al. 2008). Isolates were selected to minimize incompatible reactions and maximize aggressiveness.

**Disease Assays**

Partial resistance to *P. sojae* in the 95 soybean cultivars was analyzed by a layer test disease assay (Dorrance et al, 2008) with two isolates, C2S1 and OH121. Briefly, the *P. sojae* inoculum for the layer test was grown on lima bean agar (18 g L⁻¹) for 14 days in a glass petri dish. The colonized agar was removed from the plates and placed into a Styrofoam cup filled with approximately 8 cm of fine vermiculite measured from the bottom of the cup. The agar layer was covered with another 4 cm of vermiculite. Eight seeds from each breeding line were placed into one inoculated cup and an additional eight seeds were placed into a non-inoculated cup as a control. The experimental units, or cups, were placed in an augmented incomplete block design due to limited space on each greenhouse bench (incomplete block). A total of 16 inoculated and non-inoculated cups with eight different checks with known partial resistance were repeated twice in each incomplete block. The traits average plant height (PH), adjusted root weight (RW), adjusted shoot weight (SW), and root rot score (RRS) were measured two weeks after planting for the inoculated (I) and the non-inoculated (N) for each cup. Root rot was rated on a 1-9 scale adapted from Dorrance et al. (2008). NRRS was used to evaluate seed
quality and secondary pathogens. Cultivars with a mean NRRS > 1.5 were removed from the study. Plant height was averaged from three representative plants from each cup. Fresh weights of the roots and shoots were adjusted by dividing the weight by the total number of plants that germinated per cup. The entire experiment was repeated twice in time for each isolate.

**Statistical Analysis**

Phenotypic data other than NRRS was analyzed by the PROC MIXED (SAS 9.3, SAS Institute 163 Inc., Cary, NC, USA) procedure using the model $Y_{ijklm} = \mu + I_h + R(I)_{hj} + K(IR)_{hk} + C_l + G(C)_{lm} + IG(C)_{hlm} + IC_{hl} + \varepsilon_{ijklm}$, where $\mu$ is the overall mean, $I_h$ is effect of the $h^{th}$ isolate, $R(I)_{hj}$ is the effect of the $j^{th}$ experimental replicate for the $h^{th}$ isolate, $K(IR)_{hk}$ is the effect of the $k^{th}$ bench in the $j^{th}$ experimental replicate for the $h^{th}$ isolate, $C_l$ is the effect of the $l^{th}$ class of entry ($l = 1, 2, 3, 4, 5, 6, 7, 8, 9$ for OX 20-8, L83-570, PI 497861A, Resnik, Williams 79, Conrad, PI 398841, Sloan, and experimental line, respectively), $G(C)_{lm}$ is the effect of the $m^{th}$ genotype within class for experimental lines (genotypic variance, $\sigma_g^2$) $IG(C)_{hlm}$ is the effect of the $h^{th}$ isolate with the $m^{th}$ genotype within the $l^{th}$ class for the experimental lines (genotypic x isolate variance, $\sigma_{gi}^2$), $IC_{hl}$ is the effect of the $h^{th}$ isolate with the $l^{th}$ class entry, $\varepsilon_{ijklm}$ is the experimental error ($\sigma_e^2$). Class entry was treated as a fixed effect while all other effects were treated as random. Best linear unbiased predictors (BLUP) from the solution of $G(C)_{lm}$ represented the genetic effect for each cultivar (Stroup, 1989).

Correlation of traits was computed using Pearson’s correlation coefficient. Simple linear regressions were carried out with the lm function in the R package stats (version 3.0.2, R Foundation for Statistical Computing, [http://www.r-project.org/](http://www.r-project.org/)), which fit the BLUP values for
all four traits against the year of release. Breakpoints for piecewise linear regressions were
determined with the piecewise.linear function of the R package SiZer (Toms and Lesperance,
2003). Confidence intervals were derived from 1000 bootstrap samplings. When 95%
confidence intervals for the breakpoint encompassed >75% of the years, the piecewise linear
regression was considered not robust and therefore not shown. The Akaike information
criterion (AIC) was employed to compare linear with piecewise linear models to identify the
model most probable to minimize information loss (Burnham and Anderson, 2002).

Results

Overall disease development in the 95 public soybean cultivars was quantitatively
distributed for the traits IRRS, IRW, ISW, and IPH (Figure 2.1). Means for inoculated root rot
scores ranged from 1 to 9, where the population mean for IRRS was 3.41. Five cultivars,
‘Burlison’, ‘KS4694’, ‘Stressland’, ‘Jack’, and ‘Pana’, had a mean IRRS less than 1.5 for at least
one isolate, indicating a possible Rps-mediated resistance response given by a mean IRRS <1.5 for
either isolate. The check cultivar performed largely as expected. Susceptible checks OX 20-
Sloan had mean IRRS of 5.2 and 4.3, respectively. Moderately susceptible check Resnik had a
IRRS of 3.3. Moderately resistant checks L83-570 and Williams 79 had mean IRRS of 1.8 and 2.5,
respectively. Resistant checks PI 398841, PI 407861A and Conrad had mean IRRS of 2.5, 2.3, and
2.8, respectively. Cultivar checks L83-570 (Rps3a), Williams 79 (Rps1c), and Resnik (Rps1k)
carried different Rps gene. As C2S1 is avirulent against Rps1c, the mean IRRS for Williams 79
may be dependent upon the virulence profile of the two P. sojae isolates used in this study
(Kaitany et al., 2001; Dorrance et al., 2003b).
There was a significant positive correlation between the traits for IRW, ISW, and IPH, and a negative correlation between IRRS and the other three traits (Figure 2.1). This is as expected because, in contrast to the other measurements of resistance, lower values for IRRS indicate greater resistance. Similar to the inoculated traits, NRW, NSW, and NPH had a significant positive correlation. In addition, most ISW and IPH traits were significantly correlated with all of the non-inoculated traits, indicating that measurements of resistance from ISW and IPH may be confounded with overall seedling vigor (Figure 2.1).

There was a significant positive regression coefficient for the response variables IRW and IPH regressed against release year. The regression coefficient for IRRS was significantly negative (Figure 2.2). Although there are significant correlations between inoculated and non-inoculated traits, the regression of non-inoculated traits against year of cultivar do not follow the same trends. NRW and NPH did not change significantly in response to year of cultivar release and, in contrast to ISW, NSW had a negative regression coefficient (Figure 2.3).

In addition to a linear model, data were fit with a piecewise linear model. Breakpoints were estimated to be 1971 and 1984 for IRW and ISW, respectively. Of the seven traits, only IRW and ISW had a robustly supported breakpoint with 95% confidence intervals spanning around 25 years (Figure 2.2). In contrast, the confidence intervals for the breakpoints of the remaining traits encompassed over 80 years and the corresponding piecewise linear models were therefore discarded. For both IRW and ISW regressed against year of cultivar release, the piecewise linear model was more probable than the linear model (Figure 2.2). Both models indicate a positive regression coefficient prior to the breakpoint year and a negative regression coefficient after the breakpoint year.
Discussion

The significant positive regression of IRW and IPH and the significant negative regression of IRRS against release year indicate clearly that public soybean cultivars (MG II and III) exhibit evidence of genetic gains for partial resistance against *P. sojae* over 80 years of breeding (Figure 2.2). A piecewise regression with a breakpoint in the 1970s or 1980s was a better fit for the IRW and ISW traits as compared to a linear model (Figure 2.2).

Although *P. sojae* was not discovered until the late 1940’s and early 1950’s (Schmitthenner, 1985), the piecewise linear regression for IRW and ISW, as well as the linear regressions for IRRS and IPH, indicate relatively consistent improvement of partial resistance to *P. sojae* in N. American soybean cultivars prior to this date. There are multiple scenarios which can explain these findings. It is plausible that *P. sojae* was present in environments prior to the formal identification of Phytophthora root rot. Therefore, selection for increased yield in environments containing *P. sojae* may have indirectly selected for increased partial resistance. Selection for resistance alleles towards other pests or pathogens may have had pleiotropic effects on partial resistance to *P. sojae*. Several quantitative disease resistance loci (QDL) for partial resistance against *P. sojae* have been found to co-localize with resistance to other pathogens. Additionally, QDL have been shown to co-localize with clusters of resistance (*R*) genes for a variety of pests and pathogens (Lee et al., 2013a, Lee et al., 2013b; Chapter 4), making it plausible that QDL for *P. sojae* could be linked to selected *R*-genes for other pests or pathogens. For example, a QDL for *P. sojae* on chromosome 13 co-localizes with *Rps*3 and *Rps*8, but also with *Rsv*1, *Rpg*1, *Rpv*1, and *Rag*2 for resistance to soybean mosaic virus, bacterial blight caused by *Pseudomonas savastanoi* pv. *glycinea*, peanut mottle virus, and soybean aphid (*Aphis glycines*), respectively (Lee et al., 2013b). Finally, selection for developmental, morphological, or
architectural traits that result in improved agronomics may pleiotropically improve partial resistance (St. Clair, 2010). For example, stem termination and maturity date can affect soybean infection by *Phomopsis longicolla* (Thomison et al., 1990). Indeed, developmental and architectural traits have changed in soybean as a result of breeding efforts since the 1920s (Rincker et al., 2014; Suhre et al., 2014) and in this study we observed a decrease in seedling shoot weight (Figure 2.3).

In the 1970s to 1980s, partial resistance to *P. sojae* may have leveled off or declined as observed in the piecewise regression for IRW and ISW. While this could indicate fixation of alleles contributing to partial resistance in breeding programs, this period of a potential plateau was also coincident with or directly following several major events in soybean breeding and production which could influence genetic gains for partial resistance. *Rps1a* was widely released in cultivars like Harosoy in 1963 (Bernard, 1964) and most soybean acreage possessed *Rps1a* by 1972 (Schmitthenner, 1985). *P. sojae* races 2 through 24 were identified in 1965 to 1984, prompting the identification of new *Rps* genes and the development of cultivars with additional or alternative *Rps* genes/alleles to *Rps1a* (Schmitthenner, 1985). Thus, in the 1970s and 1980s, breeders may have focused on improving soybean cultivars with specific *Rps* genes. *Rps* genes can mask partial resistance, unless concerted efforts are made to select a *P. sojae* isolate which is virulent on any *Rps* gene present in the breeding material. Therefore, genetic gains for partial resistance may have declined during this time period when disease assays may have been specific for selection of novel *Rps* genes.

In addition to coinciding with our identification of diverse races of *P. sojae*, the period in which genetic gains for partial resistance may have leveled off or declined was also directly following the Plant Variety Protection (PVP) Act of 1970. Rincker et al. (2014) reported an
increase in the rate of genetic gains for yield post 1965 in a set of historic cultivars which included the cultivars assessed here as well as cultivars from private industry. This increase was attributed to a greater investment in commercial and public soybean breeding resulting from the PVP Act. It should be noted that industry released cultivars were not included in this study. Therefore, the trends reported here only represent public cultivars. Targets of selection may have differed in industry breeding programs and it would be interesting to compare partial resistance to \textit{P. sojae} between industry and public breeding programs for any differences in genetic gains, as well as to observe if genetic gains plateaued in the 1970s to 1980s.

The results from this study indicate that there have been significant changes over time for partial resistance in 80 years of soybean breeding in the U.S. Because genetic gains of partial resistance began prior to the discovery of \textit{P. sojae}, this may be attributed to indirect selection for other traits, such as yield or other disease resistance genes. There were also some factors indicating a decline in the rate of genetic gains for partial resistance, possibly due to an increased focus on race-specific resistance as virulent \textit{P. sojae} races became prevalent in the early 1970s. The continuous evolution of \textit{P. sojae} populations makes partial resistance an important trait in soybean breeding, where efforts should be made to ensure that genetic gains for partial resistance to \textit{P. sojae} continues to progress.
Figure 2.1. Histogram (diagonal), scatterplot (lower), and correlation coefficients (upper) for the four inoculated and three non-inoculated traits. The inoculated traits are labeled as followed: IRRS: Root Rot Score, IRW: Root Weight, ISW: Shoot Weight, IPH: Plant Height. The non-inoculated traits are labeled as followed: NRW: Root weight, NSW: Shoot Weight, NPH: Plant Height. *, **, *** indicates significance at the 0.05, 0.01, and 0.001 levels. Numbers in histograms indicate checks (1 = Conrad, 2 = L83-570, 3=OX-20, 4=PI398841, 5=PI407861A, 6=Resnik, 7=Sloan, and 8=Will79) with the following exceptions for checks which were off this scale of resistance or susceptibility: Check 2 is not indicated on the histogram for all traits, except root rot score. Check 3 is not indicated on the histogram for root rot score or non-inoculated root weight. Check 5 is not indicated on the histogram for non-inoculated plant height. Check 6 is not indicated on the histogram for non-inoculated shoot weight. Check 7 is not indicated on the histogram for all traits, except for non-inoculated Plant Height.
Figure 2.2. Simple linear regression (solid blue line) and piecewise linear regression (dashed line) for BLUP values of IRRS (A), IRW (B), ISW (C), and IPH (D) against the year of release corresponding to the soybean cultivar (RelYr). Breakpoint in piecewise linear regression is indicated by a vertical line. Grey shaded regions represent 95% confidence intervals for the breakpoint year. Confidence intervals for IPH and RRS breakpoint years spanned >75% of the release years, therefore no piecewise linear regression is shown.
† Significance of regression coefficient (t-test).
Figure 2.3. Simple linear regressions (solid blue line) for BLUP values of NRW (A), NSW (B), and NPH (C) against the year of release corresponding to the soybean cultivar. In piecewise linear models, confidence intervals for NRW, NSW, and NPH spanned >75% of the release years, therefore no piecewise linear regression is shown.

† Significance of regression coefficient (t-test).
Chapter 3

Genome-Wide Association Mapping and Genomic Selection for Partial Resistance to *Phytophthora sojae* in a Soybean Breeding Population

Abstract

*Phytophthora sojae* Kaufmann and Gerdemann is a destructive oomycete pathogen of soybean (*Glycine max* (L.) Merr), which causes a negative impact on soybean yields throughout the US. Genetic resistance is the recommended method for managing this pathogen, with resistance controlled in two distinct manners, *Rps* mediated resistance and partial resistance. Unlike *Rps*-mediated resistance that is race specific, partial resistance is quantitatively inherited, non-race specific, and is theoretically durable. Resistance to *P. sojae* was evaluated for 271 breeding lines with two isolates of *P. sojae*. Genome-wide association (GWA) mapping of partial resistance was conducted on a subset of 221 nonimmune lines and failed to identify any significant marker-trait associations. In contrast, genomic selection (GS) displayed the capacity to account for small-effect QTL in this highly quantitative trait. The relative efficiency of GS compared to phenotypic selection ranged from 0.5 to 1. Multi-trait selection indices allowed for the integration of multiple measures of disease and yield. The base and heritability × base selection indices performed similarly, with higher prediction accuracies than the Smith-Hazel indices. GS was shown to be a potentially highly effective method to improve partial resistance to *P. sojae* in this population.
Introduction

Soybean [Glycine max (L.) Merr] is an important economic crop in the US, setting a record high of $28 billion dollars in soybean exports in 2012-2013 (United Soybean Board, 2013). The large contribution of soybean to the US economy makes it essential to protect soybean yields. Phytophthora root and stem rot is one of the most destructive diseases of soybean (Dorrance et al., 2003a), causing yield suppression across most regions of the US (Wrather and Koenning, 2006). The causal pathogen is the soil-borne oomycete Phytophthora sojae, which thrives in wet, poorly-drained soils with the ability to infect soybeans at any given growth stage (Schmitthenner, 1985).

Management of Phytophthora sojae is primarily through single gene resistance conferred by resistance to P. sojae (Rps) genes. Rps genes provide a race-specific resistance through a gene-for-gene interaction with the pathogen’s Avirulence (Avr) genes (Flor, 1955) and confer an immune response following infection with an avirulent race of P. sojae. Consequently, Rps genes place a strong selection pressure on the pathogen population, whereby the population can evolve to overcome the Rps gene. P. sojae populations have adapted to deployed Rps genes and field populations are often composed of highly diverse pathotypes (Dorrance et al., 2003b, Grau et al., 2004). Dependent on inoculum density and environmental conditions, Rps genes have had a useful lifespan of 8 to 15 years once they have been widely deployed (Schmitthenner, 1985).

In contrast to resistance mediated by Rps genes, partial resistance, or quantitative resistance, is theoretically non-race specific and is considered to be a more durable resistance against this aggressive pathogen (Schmitthenner, 1985; Tooley and Grau, 1984). Stability of partial resistance can be attributed to its’ quantitative inheritance, where multiple genes each
contribute a small portion of genetic variance. High levels of partial resistance are often
phenotypically characterized by reduced disease symptoms, such as a reduction in the number
of successful infections, the amount of pathogen colonization and reproduction, and an increase
in the latent period of infection (Van der Plank, 1968; Young, 1996; Umaerus, 1970; Mideros et
al., 2007; Poland et al., 2009; Kou and Wang, 2010; St. Clair, 2010). Thus, increasing partial
resistance in soybean cultivars can slow the development and evolution of pathogen
populations.

Although cultivars developed in Ohio generally have high levels of partial resistance (St.
Martin et al., 1997; St. Martin et al., 2008; Mian et al., 2008), little is known about the genetic
regions associated with partial resistance to P. sojae as genetic studies have been completed in
very few N. American soybean cultivars. While a number of quantitative trait loci (QTL) have
been identified from plant introductions originating from China, Japan, or the Republic of Korea
(Nguyen et al., 2012; Lee et al., 2013a; Lee et al., 2013b; Lee et al., 2014), thus far, QTL studies in
N. American germplasm have largely been limited to resistance contributed from the cultivar,
‘Conrad’ (Burnham et al., 2003; Weng et al., 2007; Han et al., 2008; Li et al., 2010; Wang et al.,
2010; Wang H., et al., 2012). Identifying QTL regions in elite Ohio breeding lines will allow for a
more informed approach of introgressing new sources of partial resistance from diverse
soybean plant introductions and could ultimately aid in the further improvement of partial
resistance in soybean cultivars.

Discovered QTL for partial resistance to P. sojae have been identified through linkage
mapping in bi-parental populations (Burnham et al., 2003; Weng et al., 2007; Han et al., 2008; Li
et al., 2010; Wang et al., 2010; Wang, H., et al., 2012; Lee et al., 2013a,b, 2014). However,
genome-wide association (GWA) mapping is an alternative approach for finding significant
marker-trait associations in natural populations and has been recently used as a tool for marker-
assisted selection (Varshney et al., 2012). To date, GWAS has been implemented in breeding
populations of numerous crop species, including barley (*Hordeum vulgare*) (Wang, M. et al.,
2012), maize (*Zea mays*) (Zhang et al., 2013), sugar beets (*Beta vulgaris*) (Wurschum et al.,
2011), and soybean (Mamidi et al., 2014; Elmer et al., 2015), which identified novel QTL for
morphological and agronomic traits of barley, flooding tolerance, sugar yield traits, iron
deficiency chlorosis and resistance to Sclerotinia (white mold), respectively. These studies show
that GWA mapping implemented in a breeding program can identify loci for direct utilization for
crop improvement.

While GWA mapping may provide targets for marker assisted selection of large effect
loci with common alleles in the population, it is expected to be relatively uninformative for the
identification of rare alleles or loci with small effect. For traits controlled by many loci of small
effect, a genomic selection (GS) approach in which markers distributed over the genome are
used to estimate the breeding value of an individual, regardless of the significance of the
marker-trait association and thereby capturing numerous small-effect loci, may be more
effective than MAS (Meuwissen et al. 2001). In a two-step process, GS first uses a “training
population” to develop a model for genomic estimated breeding values (GEBVs) by fitting
genotypic and phenotypic data (Jannik et al 2010). In a second step, the model is applied to
genotypic data collected from an experimental population and selections for breeding
advancement are based on the GEBVs. A cross-validation approach can be utilized to validate
the GS model prior to implementation of GS (Desta and Ortiz, 2014). Simulation and empirical
studies have shown GS to outperform phenotypic selection per unit time by accelerating the
breeding cycle (Lorenzana and Bernardo, 2009) and to outperform MAS through greater genetic gains (Bernardo and Yu, 2007).

Assays for partial resistance to *P. sojae* are both labor and time intensive. Thus, GS or GWA mapping in combination with MAS can be powerful molecular tools for enhancing partial resistance to *P. sojae* in N. American soybean elite breeding lines. The first objective in this study was to identify QTL associated with resistance to *P. sojae* in a N. American soybean population using GWA mapping. The second objective was to determine the prediction accuracy of GS and compare GS relative to phenotypic selection.

**Materials and Methods**

*Plant Material*

The breeding population consisted of 271 *F*<sub>4</sub>-derived with maturity groups (MG) II and III lines from the Ohio State University soybean breeding program. Breeding lines were grown and harvested from the OARDC-OSU research station in Wooster, OH in 2012. Eight cultivar checks with known levels of partial resistance were included in the phenotypic disease assay as part of the experimental design. All of the seed was surface sterilized following a chlorine gas protocol adopted from Olhoft et al. (2006).

*Genotypic Data*

Young leaf tissue was collected, pooled, and preserved in liquid nitrogen in a 96 well plate from the first true leaves of 20 seedlings of each breeding line, grown in the Howlett greenhouse (Columbus, OH). DNA was isolated according to Keim et al. (1988).
SNP genotyping was performed with the Illumina Infinium BARCSoySNP6K BeadChip at the University of California, Davis Genome Center. These markers were selected for even genome distribution and high polymorphism information content (Song et al., 2013). Marker genotypes were called using GenomeStudio software version 2011.1 (Illumina Inc., San Diego, CA). Monomorphic markers, markers with >5% missing data, and markers with >25% missing plus heterozygous alleles were removed from the dataset. Missing data was imputed with fastPHASE (Scheet and Stevens, 2006).

Collection of Phenotypic Data

Isolates C2S1 and OH12108_6.3 (OH121) were selected for virulence and aggressiveness as described in Chapter 2. For the elimination of a race-specific response, or incompatible interaction, two pathogenic tests were conducted on 27 isolates for virulence and aggressiveness. The hypocotyl inoculation test, adopted from Dorrance et al. (2008), evaluates pathogen virulence by the percentage of susceptible (compatible reaction), and resistant (incompatible reaction) plants for fifteen soybean differentials. The aggressiveness of the isolates were measured in the tray test (Olah and Schmitthenner, 1985), through the mean lesion length of soybean taproots in ten seedlings in the cultivars, ‘Sloan’ and ‘Conrad’ with moderate susceptibility and high level of partial resistance, respectively (Burnham et al. 2003). Isolates OH12108_6.3 (OH121) and C2S1 were identified as a moderately aggressive isolate with high virulence (vir 1a, 1b, 1c, 1d, 1k, 2, 3a, 3c, 4, 5, 6, 7, 8 and vir 1a, 1b, 1d, 1k, 2, 3a, 3b, 3c, 4, 5, 6, 7, 8, respectively) eliminating race-specific response in most of the differential lines (Table 3.1; Charlotte Smith, 2014).
Evaluation of partial resistance to *P. sojae* in 271 breeding lines with the two selected isolates was carried out with the layer test (Dorrance et al., 2008) as described in Chapter 2. Briefly, *P. sojae* colonized lima bean agar was placed into a Styrofoam cup filled with approximately 8 cm of fine vermiculite measured from the bottom of the cup. Eight seeds from each breeding line was placed into one inoculated cup and additional eight seeds are placed into a non-inoculated cup as a control. The experimental units, or cups, were organized in an augmented incomplete block design due to limited space in the greenhouse. A total of 16 inoculated and non-inoculated cups with eight different checks with known partial resistance were repeated twice on each greenhouse bench, or each incomplete block. Average plant height taken from three seedlings per cup (PH), adjusted root weight (RW), adjusted shoot weight (SW), and root rot score (RRS) were measured two weeks after planting for the inoculated (I) and the non-inoculated (N). Root rot was rated on a 1-9 scale adapted from Dorrance et al. (2008). Fresh weights of the roots and shoots were adjusted by dividing by the total number of plants per cup. The experiment was repeated twice in time for each isolate. Lines with NRRS > 1.5 were removed from the dataset, eliminating possible variation caused by contamination or seedborne diseases.

Breeding lines were grown in 2012 for collection of yield data. Lines were grown in separate yield tests for early or late maturity (MGII or MGIII) and preliminary or advanced lines depending on their maturity group and their generation (F4:6 or F4:7-9), respectively. Lines for each test were planted in a randomized complete block design with two replicates for the preliminary lines (F4:6) or three replicates for the advanced lines (F4:7-9) at three Ohio locations, South Charleston, Plain City, and Wooster in 2012. The early maturing advanced lines, the late maturing advanced lines, and the specialty lines were planted in six row plots with a
The preliminary tests for the early and late maturing lines were planted in South Charleston in three row plots with a plot size of 4 m x 1.3 m. In Hoytville and Wooster, the preliminary tests were planted in two row plots with a plot size of 4.9 m x 0.6 m. Seeds were harvested from each location during the fall of 2012. Check cultivars were included in multiple yield tests and used to normalize data.

**Phenotypic data analysis**

While highly virulent isolates were selected for the disease assays in order to avoid *Rps* mediated resistance, some breeding lines possessed an average root rot score of less than 1.5 in layer test assays with either *C2S1* or *OH121*. Because we could not determine if this lack of disease progress was due to an *Rps*-mediated resistance or high levels of partial resistance, the data was grouped into two sets. The full dataset consisted of all 271 breeding lines, the nonimmune data set consisted of 224 breeding lines with an average IRRS greater than or equal to 1.5 for each isolate.

Analysis of PH, RW, SW, and RRS was carried out with the PROC MIXED procedure (SAS 9.3, SAS Institute 163 Inc., Cary, NC, USA). Data was analyzed with two models. Using observations from both isolates, the model included interactions with isolate: 

\[ Y_{ijkm} = \mu + I_h + R(I)_{hj} + K(IR)_{hjk} + C_l + G(C)_{lm} + IG(C)_{hlm} + IC_{hl} + \varepsilon_{hjklm}, \]

where \( \mu \) is the overall mean, \( I_h \) is effect of the \( h \)th isolate, \( R(I)_{hj} \) is the effect of the \( j \)th experimental replicate for the \( h \)th isolate, \( K(IR)_{hjk} \) is the effect of the \( k \)th bench in the \( j \)th experimental replicate for the \( h \)th isolate, \( C_l \) is the effect of the \( l \)th class of entry (\( l = 1, 2, 3, 4, 5, 6, 7, 8, 9 \) for OX 20-8, L83-570, PI 497861A, Resnik, Williams 79, Conrad, PI 398841, Sloan, and experimental line, respectively), \( G(C)_{lm} \) is the effect of the \( m \)th genotype within class for experimental lines (genotypic variance, \( \sigma_g^2 \)), \( IG(C)_{hlm} \) is the effect of the \( h \)th isolate...
with the $m^{th}$ genotype within the $l^{th}$ class for the experimental lines (genotypic x isolate variance, $\sigma^2_{gi}$), $IC_{il}$ is the effect of the $h^{th}$ isolate with the $l^{th}$ class entry, $\epsilon_{hjk}$ is the experimental error ($\sigma^2$).

Broad-sense heritability ($H^2$) was calculated for each trait as follows: $H^2 = \frac{\sigma^2_g}{\sigma^2_g + \sigma^2_gi/r}$, where $r$ is the number of experimental replicates and $i$ is the number of isolates.

Using observation with one isolate at time, the model did not include isolate or interactions with isolate: $Y_{ijklm} = \mu + R_j + K(R)_{jk} + C_l + G(C)_{lm} + \epsilon_{hjk}$. For each model, best linear unbiased predictors (BLUP) from the solutions of $G(C)_{lm}$ represented the genetic effect for each cultivar (Stroup, 1989). Broad-sense heritability ($H^2$) was calculated for each trait as follows: $H^2 = \frac{\sigma^2_g}{\sigma^2_g + \sigma^2_gi/r}$.

Correlations of BLUP values for traits were computed using Pearson’s correlation coefficient implemented in R (version 3.0.2, R Foundation for Statistical Computing, http://www.r-project.org/).

BLUPs were also calculated for yield using the model: $Y_{lnmo} = \mu + C_l + G_m(C_l) + L_m + T_n(L_m) + L_m \times G_m(C_l) + L_m \times G_m + \epsilon_{lnmo}$, where $L_m$ is the effect of the $m^{th}$ location, $T_n(L_m)$ is the effect of the $n^{th}$ trial within the $m^{th}$ location, $L_m \times G_m(C_l)$ is the interaction effect of the $n^{th}$ location with the $m^{th}$ cultivar or breeding line within a class, $L_m \times C_l$ is the interaction effect of the $n^{th}$ location with the $l^{th}$ class and other variables are as described above. Class was treated as a fixed effect, all other effects was treated as random.

Three different selection indices were calculated for partial resistance to $P. sojae$ which included IRRS, IRW, ISW, and IPH. Selection indices for yield plus partial resistance to $P. sojae$ included IRRS, IRW, ISW, and IPH and yield. Weights for the base selection index (Williams, 1962) were derived from the correlation of IRRS, IRW, ISW, and IPH with field yields under diseased conditions (Wang, H., et al. 2012). Yield was assigned a weight four times greater than
the sum of the absolute values of the inoculated trait weights, such that the inoculated traits were equal 20% and yield was 80% of the index. The $H^2 \times \text{base}$ index multiplied the base weights for each trait by their heritabilities. The Smith-Hazel selection index (Smith, 1936; Hazel, 1943) was calculated by multiplying the inverse of the phenotypic covariance matrix by the covariance matrix of genetic effects and the base index weights for each trait. The phenotypic covariance matrix was derived from raw trait values, whereas the covariance matrix of genetic effects was derived from BLUP values.

**GWA Mapping**

Analyses of the phenotypic and genotypic data sets for the full dataset of 271 breeding lines and the 224 nonimmune breeding lines were conducted with GAPIT (Genome Association Prediction Tool) (Lipka et al., 2012) in R software implemented with the compressed mixed linear model (CMLM) with population parameters previously determined and the number of principle components used as covariates determined by Bayesian Information Criterion (Zhang et al., 2010). Similar to the mixed linear model (MLM) (Yu et al., 2006), CMLM can be described as $Y = X\alpha + P\beta + Zu + e$, where $Y$ is the phenotypic score, $X$ is the genotypic scores, $P$ is population structure (PCA matrix), $Z$ is the relatedness of individuals (Kinship matrix), and $e$ is the residual due to error (Huang et al., 2010). The fixed effects $X\alpha$ and $P\beta$ represent the distant relatedness within the population. The random effect $Zu$ represents the recent shared ancestry between the individuals in the population and is a random effect. The significance threshold for marker-trait associations was determined by a modified Bonferroni adjustment in which the effective number of independent tests ($M_{\text{eff}}$) was calculated using simpleM (Gao et al., 2008).
Genome-wide threshold levels for each of the two datasets were determined by $\alpha/M_{eff}$, where $\alpha=0.05$ (Miller, 1981).

**Genomic Selection**

Genomic selection (GS) was conducted on the nonimmune breeding lines using the same phenotypic and genotypic data used in the GWAS. MLM was applied to the training population data using three methods of compression: no compression (MLM; Yu et al., 2006), compression (CMLM; Zhang et al., 2010), and enriched compression (ECMLM, Li et al., 2014), all implemented in the GAPIT package. The training population was randomly sampled from the nonimmune breeding lines and consisted of 90% (202) of the nonimmune breeding lines. Genomic estimated breeding values (GEBVs) were calculated for the inference panel consisting of the remaining 10% (22) nonimmune breeding lines. Prediction accuracies ($r_{pa}$) for genomic selection were calculated as the correlation between the GEBV and observed values for the inference panel. One hundred iterations were carried out, a single factor ANOVA on the GEBVs for each trait determined significant differences between the three compression levels.

In addition to predicting genotypes across a single environment, $r_{pa}$ across environments or isolates was determined by correlating GEBVs from one isolate to the phenotypic observations of the second isolate. The relative efficiency of GS as compared to phenotypic selection was calculated for each trait as the ratio between the across isolate $r_{pa}$ and the phenotypic correlation between isolates ($r_p$).
 Results

Phenotypic Data Analysis

Based on a comparison of BLUP values for IRRS to the mean solution for estimated fixed effects of the resistant cultivar checks, L83-570, Williams 79, PI 407861A, PI 398841, and Conrad, 81% of the 271 OSU breeding lines were more resistant than checks. Sixteen percent of OSU breeding lines were more susceptible that the check cultivars Sloan and OX 20-8. All measurements of resistance were significantly correlated with each other ($p$-value < 0.001). The BLUP values for IRW, ISW, and IPH were positively correlated with each other and BLUP values for IRRS were negatively correlated with the other traits (Figure 3.1). Genetic variation was significant for traits in both the full dataset, which included immune lines with IRRS <1.5, and the nonimmune dataset in an analysis of phenotypes from both isolates with a model which included a genotype x isolate interactions as well as for isolate OH121 alone. However, there was no estimable genetic variation for IRRS as observed from isolate C2S1 alone and genetic variation for IRW and IPH from C2S1 alone were insignificant (Table 3.2). $H^2$ of traits with C2S1 alone was none to moderate (0 to 0.438). Whereas $H^2$ for traits analyzed from both isolates or OH121 alone were moderate to high (0.405 to 0.808). The isolate C2S1 exhibited less disease development as compared to OH121 (Figure 3.2). For each inoculated trait, the distribution of raw values was narrower and lower on the disease scale for assays with C2S1 as compared to assays with OH121, this is especially apparent for IRRS (Figure 3.2).

NRW, NSW, and NPH also had high heritability (0.696 to 0.764) (Table 3.2). In comparisons between the full dataset and the nonimmune dataset, the full dataset had higher heritability versus the nonimmune dataset for all inoculated traits; however, heritabilities were similar for non-inoculated traits between the nonimmune and full dataset.
**GWA Mapping**

A total of 3,113 polymorphic markers were used in GWA mapping. There were no significant marker-trait associations in nonimmune datasets for inoculated traits (Figure 3.3). To identify potential novel Rps genes, the full dataset was also assessed; however, no significant marker-trait associations were identified (data not shown). The most significant markers, ss715631579 (p-value = 2.85E-04) and ss715631594 (p-value= 3.24E-04), were located on chromosome 18:50325785 and 18:50504231 and were close to the genome-wide p-value threshold of 2.80E-05, for both IRW and ISW in the full dataset (Figures 3.2A and 3.2B). No markers on chromosome 18 were near significance for the non-inoculated traits, however a significant marker, ss715622631, for NRW was observed on chromosome 15: 50359609 (Figures 3.2C and 3.2D). No significant marker-trait associations were found for any of the four inoculated traits which were isolate specific to C2S1 or OH121 (data not shown).

**Genomic Selection**

Regardless of the method of compression, there were few significant differences among MLM, CMLM, or ECMLM for the inoculated and noninoculated traits. Exceptions were that MLM was significantly better than CMLM for yield and than CMLM or ECMLM for the $H^2\times$base index for Yield with Partial resistance, ECMLM was significantly worse than MLM or CMLM for the Smith-Hazel index for Partial resistance (Table 3.3). While generally not significantly different, the MLM had highest prediction accuracies for eight out of 14 of the tests with average means ranging from 0.229 to 0.470 compared to the CMLM method with an average
mean ranging from 0.168 to 0.449 and the ECMLM method had average means ranging from 0.173 to 0.444 (Table 3.3).

The base selection index and $H^2 \times$ base selection index performed similarly (Table 3.3). The $r_{pa}$ of the Smith-Hazel index was significantly lower than that of either the base selection index or the $H^2 \times$ base selection index, the one exception to this was in the application of the ECMLM to multi-trait indices for Yield with Partial resistance (Table 3.3).

The prediction accuracy differed significantly when data from $OH121$ isolate was used in training the model and data from $C2S1$ was used as the inference panel as compared to $C2S1$ as the training data and $OH121$ as the inference data (Table 3.4). However, which isolate was best used in training and inference varied depending on the trait. Use of $C2S1$ as the training data and $OH121$ as the inference panel resulted in significantly higher prediction accuracies than when data from $OH121$ assays were used to predict $C2S1$ for IRW (Table 3.4). The reverse was true for IPH, using $OH121$ as the training data and $C2S1$ as the inference panel resulted in significantly higher prediction accuracies (Table 3.4). Prediction accuracy was significantly higher when data from $C2S1$ assays were used in training for the selection indices. The phenotypic correlation between $C2S1$ and $OH121$ ($r_p$) ranged from 0.301 to 0.376 for the four inoculated traits and 0.311 to 0.330 for the selection indices for Partial resistance. Thus, the efficiency of genomic selection to phenotypic selection across isolates ($r_{pa}/r_p$) ranged from 0.411 to 1.042 (Table 3.4).
Discussion

Limitations in detecting significant marker-trait associations

This is the first study conducting GWA and GS for partial resistance to *P. sojae* in a breeding population. Based on the lack of significant marker-trait associations, it is presumed that partial resistance to *P. sojae* in these breeding lines is either conferred by rare alleles or by many genes of small effect. Previous QTL studies of a bi-parental population between two cultivars, Conrad and Sloan, reported five small effect QTL explaining 4-7% phenotypic variation (Wang et al., 2010) or three small effect QTL explaining less than 5% and one QTL explaining 22.6% of the phenotypic variation (Wang, H., et al., 2012) add support to partial resistance of *P. sojae* being controlled by many genes of small effect. The most significant markers, ss715631579 and ss715631594, located on chromosome 18 are approximately 8-9 Mb away from a single QTL associated with partial resistance to *P. sojae* conducted in a tray test for the trait, mean lesion length (Wang H et al., 2012). Importantly, no significance or near significance were found for NRS, NSW, or NPH on chromosome 18, thus, it is likely that these markers are contributing to partial resistance and not simply to overall seedling vigor.

The lack of significant marker-trait associations limits the ability to utilize marker-assisted selection for improvement of partial resistance in this population (Zhong et al., 2009) and highlight the value of genomic selection as a means of applying MAS to a highly quantitative trait. In a barley study of MAS for Fusarium Head Blight (FHB) Resistance and deoxynivalenol (DON) accumulation, MAS using a subset of significant markers was ineffective; yet, GS achieved prediction accuracies of 0.72 for FHB resistance and 0.68 for DON concentration (Lorenz et al., 2012). Although no significant markers were identified in the present research to compare
MAS to GS, further assessments could be made to compare GS to MAS using a limited number of markers with highest significance.

*Prediction accuracies and relative efficiencies of GS*

Utilizing the same environment in the training population and inference panel for cross validation of GS can result in upward bias of prediction accuracy, while using an inference panel environment which is absent from the training population is preferable for the calculation of prediction accuracies and may approximate real-world scenarios. Disease phenotypes used in these studies were from greenhouse trials, and are therefore dissimilar to real-world field environments. However, the utilization of multiple isolates allowed us to assess how a model developed with one isolate would predict the phenotypes of individuals inoculated with another isolate. Prediction accuracies measured across isolates (Table 3.4) were lower than prediction accuracies estimated from the same training and inference environments (Table 3.3), which had an average 38% upward bias across all traits and selection indices. The relative efficiency of GS to phenotypic selection ranged from 0.411 to 1.042, and compared favorably to studies of grain dry matter in maize across locations and years where the relative efficiencies were higher in environment 1 (0.58) than environment 2 (0.38) (Albrecht et al., 2014) and were similar to those reported for agronomic and grain quality in a wheat population where relative efficiencies ranged from 0.84 to 1.08 (Heffner et al., 2011).

The prediction accuracy of a GS model is partly dependent on the model, the heritability of the trait, and the distribution of markers such that each gene is in linkage disequilibrium with at least one marker (Heffner et al., 2011; Burgueno et al., 2012; Hickey et al., 2012). Ridge regression (rrBLUP) which assumes all markers are random effects with a common variance and
Gaussian distribution is a widely used genomic selection method for traits in which there are no QTL of large effect (Resende et al., 2012; Schulz-Streeck et al., 2012; Desta and Ortiz, 2014).

Without compression, GEBVs from MLM applied in GAPIT are equivalent to rrBLUP (vanRaden et al., 2008). While compression and enhanced compression have been shown to increase power relative to no compression (MLM) in GWA mapping (Yu et al., 2006; Zhang et al., 2010; Li et al., 2014), no systematic comparison of models developed with and without compression has been done for GS. In this study, models developed with and without compression or enhanced compression generally resulted in similar prediction accuracies (Table 3.2). Though MLM, equivalent to the popular rrBLUP GS model, most frequently resulted in the highest prediction accuracy.

Prediction accuracy was not significantly correlated with heritability (P-value > 0.05). While Zhong et al. (2009) observed lower prediction accuracies of 0.56 to 0.62 when narrow-sense heritability ($h^2$) was 0.4 compared to prediction accuracies of 0.61 and 0.66 when $h^2$ was 0.75 and 0.86, this relationship was not observed in our study where there was no consistent trend between $H^2$ and prediction accuracy for NRW, NSW, NPH, IRRS, IRW, ISW, IPH, and yield. The heritabilities of several traits in the C2S1 dataset were very low (< 0.1), likely due to poor disease development and highlighting the importance of utilizing multiple isolates in disease screens and selection for aggressive isolates. Where C2S1 had very low heritability for ISW and IPH (Table 3.1), prediction accuracies utilizing OH121 as the training dataset were better than using C2S1 as the training dataset. Taken together, this confirms that heritability plays a role in prediction accuracy of a GS model and there are likely additional elements which are also critical.
The application of multi-trait indices in GS

We used four traits as estimates for disease development. Multi-trait indices can be utilized to integrate these into a single GEBV (e.g., Heffner et al., 2011). The implementation of GS for disease resistance in a breeding program will likely necessitate selection for both disease resistance and yield (Bao et al., 2014). Therefore, we evaluated multi-trait selection indices with the four measures of resistance to *P. sojae* and also included multi-trait selection indices with these measures of resistance to *P. sojae* and yield, where yield was weighted heavily. In this study, the base index and $H^2 \times \text{base index}$ resulted in similarly high prediction accuracies and relative selection efficiencies. While the base index which ignores phenotypic and genotypic correlations between traits and is theoretically inferior to the Smith-Hazel index, previous studies have reported a range of outcomes including the superiority of the Smith-Hazel index (Heffner et al., 2011), the effective equality of the two indices (Sabouri et al., 2008; Eshghi et al., 2011), as well as the superiority of the base index (Jost et al., 2012). The utility of the Smith-Hazel index is highly dependent on the accuracies of the genetic and phenotypic covariance matrices (Williams et al., 1962; Lin, 1978); thus where these matrices may be inaccurate due to sampling errors in a limited population size, the base index is preferred (Harris, 1964).

Overall, this study indicates that the implementation of GS in a population of breeding lines can effectively improve partial resistance to *P. sojae*, a highly quantitative trait controlled by many genes with small effects. Importantly, GS can be implemented through use of multi-trait indices which include yield. As work progresses towards the discovery of exotic germplasm with high levels of partial resistance (Tucker et al., 2010; Nguyen et al., 2012; Lee et al., 2013a,
b; Lee et al., 2014; Chapter 4), GS may be a valuable tool for improving the partial resistance to *P. sojae* with these sources while maintaining or improving yield.
<table>
<thead>
<tr>
<th>Isolate</th>
<th>Average Lesion Length (cm)</th>
<th>Virulence profile</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Conrad</td>
<td>Sloan</td>
</tr>
<tr>
<td>1S11</td>
<td>14.19</td>
<td>16.56</td>
</tr>
<tr>
<td>9d</td>
<td>ND†</td>
<td>ND</td>
</tr>
<tr>
<td>Ash 1-1</td>
<td>12.44</td>
<td>12.08</td>
</tr>
<tr>
<td>Butler MU S.1</td>
<td>9.13</td>
<td>15.25</td>
</tr>
<tr>
<td><strong>C2S1</strong></td>
<td><strong>12.50</strong></td>
<td>ND</td>
</tr>
<tr>
<td>DFL-1</td>
<td>5.94</td>
<td>20.19</td>
</tr>
<tr>
<td>DPS</td>
<td>16.44</td>
<td>18.63</td>
</tr>
<tr>
<td>MMI12004_04_0.2</td>
<td>18.19</td>
<td>29.94</td>
</tr>
<tr>
<td>OH1200110.1</td>
<td>15.88</td>
<td>20.69</td>
</tr>
<tr>
<td>OH120049.1</td>
<td>23.38</td>
<td>16.88</td>
</tr>
<tr>
<td>OH120484.1</td>
<td>15.78</td>
<td>ND</td>
</tr>
<tr>
<td>OH121033.4</td>
<td>15.38</td>
<td>21.94</td>
</tr>
<tr>
<td>OH121047.1</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>OH121074.4</td>
<td>10.00</td>
<td>ND</td>
</tr>
<tr>
<td><strong>OH121086.3</strong></td>
<td><strong>14.00</strong></td>
<td><strong>23.00</strong></td>
</tr>
<tr>
<td>OH121087.1</td>
<td>19.88</td>
<td>22.06</td>
</tr>
<tr>
<td>OH121108.2</td>
<td>16.69</td>
<td>22.43</td>
</tr>
<tr>
<td>OH121187.2</td>
<td>12.00</td>
<td>20.63</td>
</tr>
<tr>
<td>OH121189.1</td>
<td>20.92</td>
<td>ND</td>
</tr>
<tr>
<td>OH121215.1</td>
<td>25.56</td>
<td>21.81</td>
</tr>
<tr>
<td>OH121018.1</td>
<td>17.31</td>
<td>16.13</td>
</tr>
<tr>
<td>OH121215.1</td>
<td>29.88</td>
<td>28.26</td>
</tr>
</tbody>
</table>

**Table 3.1** Summary of mean lesion lengths from tray tests and virulence profiles from hypocotyl inoculations utilizing 23 different isolates of *P. sojae*. Isolates selected for disease assays in this study are indicated in bold.

† ND: Not determined
<table>
<thead>
<tr>
<th></th>
<th>Nonimmune</th>
<th></th>
<th>Full</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$H^2$</td>
<td>Sig. of $\sigma^2_g$ (p-value)</td>
<td>$H^2$</td>
</tr>
<tr>
<td>Non-inoculated Root Weight</td>
<td>0.758</td>
<td>&lt;.0001</td>
<td>0.764</td>
</tr>
<tr>
<td>Non-inoculated Shoot Weight</td>
<td>0.760</td>
<td>&lt;.0001</td>
<td>0.761</td>
</tr>
<tr>
<td>Non-inoculated Plant Height</td>
<td>0.696</td>
<td>&lt;.0001</td>
<td>0.718</td>
</tr>
</tbody>
</table>

Model: $Y_{ijklm} = \mu + R_j + K(R)_{jk} + C_i + G(C)_{lm} + \varepsilon_{ijklm}$
Non-inoculated

Inoculated Root Rot Score | 0.405     | <.0001               | 0.422 | <.0001 |
Inoculated Root Weight    | 0.661     | <.0001               | 0.659 | <.0001 |
Inoculated Shoot Weight   | 0.560     | <.0001               | 0.643 | <.0001 |
Inoculated Plant Height   | 0.512     | <.0001               | 0.523 | <.0001 |

Model: $Y_{ijklm} = \mu + I_h + R(I)_{jh} + K(I)_{jh} + C_i + G(C)_{lm} + IG(C)_{lhm} + IC_{hl} + \varepsilon_{ijklm}$
Isolates: C2S1 and OH121086.3

Inoculated Root Rot Score | 0.405     | <.0001               | 0.422 | <.0001 |
Inoculated Root Weight    | 0.661     | <.0001               | 0.659 | <.00001 |
Inoculated Shoot Weight   | 0.560     | <.0001               | 0.643 | <.0001 |
Inoculated Plant Height   | 0.512     | <.0001               | 0.523 | <.0001 |

Table 3.2. Heritability and genetic variance of each trait for the nonimmune and full datasets.

† Sig of $\sigma^2_g$: significance of genetic variance

‡ NE: Not estimable
<table>
<thead>
<tr>
<th>Trait†</th>
<th>Multi-trait selection index</th>
<th>Mean prediction accuracies $(r_{pa})$</th>
<th>MLM‡</th>
<th>CMLM§</th>
<th>ECMLM¶</th>
</tr>
</thead>
<tbody>
<tr>
<td>NRW</td>
<td>NA¹</td>
<td>0.229</td>
<td>0.168</td>
<td>0.173</td>
<td></td>
</tr>
<tr>
<td>NSW</td>
<td>NA</td>
<td>0.338</td>
<td>0.350</td>
<td>0.330</td>
<td></td>
</tr>
<tr>
<td>NPH</td>
<td>NA</td>
<td>0.485</td>
<td>0.499</td>
<td>0.509</td>
<td></td>
</tr>
<tr>
<td>IRRS</td>
<td>NA</td>
<td>0.421</td>
<td>0.395</td>
<td>0.410</td>
<td></td>
</tr>
<tr>
<td>IRW</td>
<td>NA</td>
<td>0.322</td>
<td>0.310</td>
<td>0.362</td>
<td></td>
</tr>
<tr>
<td>ISW</td>
<td>NA</td>
<td>0.470</td>
<td>0.449</td>
<td>0.443</td>
<td></td>
</tr>
<tr>
<td>IPH</td>
<td>NA</td>
<td>0.461</td>
<td>0.443</td>
<td>0.420</td>
<td></td>
</tr>
<tr>
<td>Yield</td>
<td>NA</td>
<td>0.331a</td>
<td>0.248b</td>
<td>0.262ab</td>
<td></td>
</tr>
<tr>
<td>Partial resistance</td>
<td>Base</td>
<td>0.450ab</td>
<td>0.395bc</td>
<td>0.418ab</td>
<td></td>
</tr>
<tr>
<td>Partial resistance</td>
<td>$H^2\times$base</td>
<td>0.420a</td>
<td>0.424ab</td>
<td>0.414abc</td>
<td></td>
</tr>
<tr>
<td>Partial resistance</td>
<td>Smith-Hazel</td>
<td>0.365cd</td>
<td>0.329d</td>
<td>0.278e</td>
<td></td>
</tr>
<tr>
<td>Yield and partial resistance</td>
<td>Base</td>
<td>0.295b</td>
<td>0.299ab</td>
<td>0.255bcd</td>
<td></td>
</tr>
<tr>
<td>Yield and partial resistance</td>
<td>$H^2\times$base</td>
<td>0.350a</td>
<td>0.272bc</td>
<td>0.220cd</td>
<td></td>
</tr>
<tr>
<td>Yield and partial resistance</td>
<td>Smith-Hazel</td>
<td>0.216d</td>
<td>0.103e</td>
<td>0.269bcd</td>
<td></td>
</tr>
</tbody>
</table>

**Table 3.3.** Mean prediction accuracies of genomic selection for plant traits and selection indices.


‡ MLM: mixed linear model

§ CMLM: compressed mixed linear model

¶ ECMLM: enhanced compressed mixed linear model

# NA: Not applicable

|| For each trait, different letters indicate significantly different prediction accuracies (Fisher’s protected t-test, $\alpha = 0.05$)
<table>
<thead>
<tr>
<th>Trait</th>
<th>Multi-trait selection index</th>
<th>Isolate for training dataset</th>
<th>Isolate for inference dataset</th>
<th>Mean prediction accuracies ($r_{pa}$)</th>
<th>Correlation of phenotypes between isolates ($r_p$)</th>
<th>Relative efficiency of GS ($r_{pa}/r_p$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRW</td>
<td>NA</td>
<td>$C2S1$</td>
<td>$OH121$</td>
<td>0.313**</td>
<td>0.301</td>
<td>1.042</td>
</tr>
<tr>
<td>IRW</td>
<td>NA</td>
<td>$OH121$</td>
<td>$C2S1$</td>
<td>0.252</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISW</td>
<td>NA</td>
<td>$C2S1$</td>
<td>$OH121$</td>
<td>0.284</td>
<td>0.356</td>
<td>0.797</td>
</tr>
<tr>
<td>ISW</td>
<td>NA</td>
<td>$OH121$</td>
<td>$C2S1$</td>
<td>0.324</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IPH</td>
<td>NA</td>
<td>$C2S1$</td>
<td>$OH121$</td>
<td>0.234***</td>
<td>0.376</td>
<td>0.623</td>
</tr>
<tr>
<td>IPH</td>
<td>NA</td>
<td>$OH121$</td>
<td>$C2S1$</td>
<td>0.340</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partial resistance</td>
<td>Base</td>
<td>$C2S1$</td>
<td>$OH121$</td>
<td>0.322a*</td>
<td>0.300</td>
<td>0.974</td>
</tr>
<tr>
<td>Partial resistance</td>
<td>Base</td>
<td>$OH121$</td>
<td>$C2S1$</td>
<td>0.179b</td>
<td>0.330</td>
<td>0.542</td>
</tr>
<tr>
<td>Partial resistance</td>
<td>$H^2 \times$base</td>
<td>$C2S1$</td>
<td>$OH121$</td>
<td>0.325a</td>
<td>0.329</td>
<td>0.987</td>
</tr>
<tr>
<td>Partial resistance</td>
<td>$H^2 \times$base</td>
<td>$OH121$</td>
<td>$C2S1$</td>
<td>0.159b</td>
<td>0.329</td>
<td>0.544</td>
</tr>
<tr>
<td>Partial resistance</td>
<td>Smith-Hazel</td>
<td>$C2S1$</td>
<td>$OH121$</td>
<td>0.205b</td>
<td>0.311</td>
<td>0.659</td>
</tr>
<tr>
<td>Partial resistance</td>
<td>Smith-Hazel</td>
<td>$OH121$</td>
<td>$C2S1$</td>
<td>0.128c</td>
<td>0.311</td>
<td>0.411</td>
</tr>
</tbody>
</table>

Table 3.4. Prediction accuracies and phenotypic correlations between isolates and relative efficiency of genomic selection (GS). For IRW, ISW, and IPH, significant differences are indicated by *, **, and *** for p-values less than 0.05, 0.01, and 0.001, respectively (F-test).


‡ NA: Not applicable

§ $OH121$: $OH121086.3$

¶ Different letters indicate significantly different prediction accuracies for Partial resistance (Fisher’s protected T-test, $\alpha = 0.05$)
Figure 3.1. Histogram (diagonal), scatterplot (lower), and Pearson’s correlation (upper) for BLUP values of the four inoculated traits. Numbers in histograms indicate checks (1 = Conrad, 2 = L83-570, 3 = OX-20, 4 = PI 398841, 5 = PI 407861A, 6 = Resnik, 7 = Sloan, and 8 = Williams 79). The BLUP values of the highly susceptible checks 3 and 7 are beyond the range of the distribution of experimental lines and are therefore not indicated in the histogram.
Figure 3.2. Overlapping histograms with the raw values of the four inoculated traits for isolates OH121086.3 and C2S1.
Figure 3.3. Manhattan plot of the soybean genome in the full data set depicting the extent of the associations of 3,113 SNPs with inoculated shoot weight (A), inoculated root weight (B), non-inoculated shoot weight (C), and non-inoculated root weight (D). The $-\log(p$-value) on the y-axis is a measure of the degree to which a SNP is associated with a trait. Chromosomes are displayed on the x-axis. The horizontal line indicates the genome-wide threshold for significance.
CHAPTER 4

Genome-Wide Association Mapping of Partial Resistance to

*Phytophthora sojae* in Soybean Plant Introductions from the Republic of Korea

Abstract

Phytophthora root and stem rot is one of the most yield-limiting diseases of soybean ([*Glycine max* (L.) Merr], caused by the oomycete *Phytophthora sojae*). Partial resistance is controlled by several genes and compared to single gene resistance to *P. sojae* controlled by *Rps* genes, places less selection pressure on the *P. sojae* populations. Thus, partial resistance provides a more durable resistance against the pathogen. Plant introductions (PIs) originating from the Republic of Korea (S. Korea) have been excellent sources for high levels of partial resistance against *P. sojae*. Resistance to two highly virulent *P. sojae* isolates was assessed in 1,398 PIs from S. Korea via a greenhouse layer test. Genome-wide association mapping was carried out on the non-immune portion of this population to identify genomic regions associated with partial resistance. Sixteen SNP markers in three regions on chromosomes 3, 13, and 19 were significantly associated with partial resistance to *P. sojae*. The coincident locations of the QTL on chromosomes 3 and 13 with *Rps* genes and genes putative encoding nucleotide binding-leucine rich repeats allow speculation of the mechanism for partial resistance conferred
at these loci. The QTL on chromosome 19 represents a novel locus for resistance to *P. sojae*. in S. Korean germplasm provide a potentially novel source for the improvement of partial resistance to *P. sojae* in soybean cultivars.
Introduction

One of the most devastating diseases of soybean \textit{[Glycine max (L.) Merr]} is Phytophthora root and stem rot, caused by a soil-borne, oomycete pathogen, \textit{Phytophthora sojae} (Kaufmann and Gerdemann), which thrives in wet soil conditions that allows their asexual, motile zoospores to colonize the cortex of the soybean root. Following the infection, symptoms in a susceptible phenotype will display a brown lesion starting from the root and moving up to the stem. This results in wilted, chlorotic leaves that will eventually kill the plant. This disease was the second most yield limiting disease between 2001 and 2010 (Wrather and Koenning, 2010).

An effective measure for controlling \textit{P. sojae} is genetic resistance. Single-gene mediated resistance to \textit{P. sojae}, or \textit{Rps} gene mediated resistance, provides complete resistance through the plant defense mechanism known as effector triggered immunity (Zhang et al., 2011). However, this resistance is race-specific and the effectiveness of a given \textit{Rps} gene is dependent on the population of \textit{P. sojae} to which the plant is exposed. Deployment of an \textit{Rps} gene places high selection pressures on the \textit{P. sojae} population causing the population to evolve and potentially become virulent such that the \textit{Rps} gene is no longer effective. Widespread deployment of \textit{Rps} genes in soybean cultivars has resulted in the evolution of highly diverse \textit{P. sojae} populations (Burnham et al., 2003). At least 55 physiological races of \textit{P. sojae} have been identified in the US, which excludes \textit{P. sojae} pathotypes with a compatible interaction to all deployed \textit{Rps} genes (Schmitthenner et al., 1994; Yang et al., 1996; Abney et al., 1997; Leitz et al., 2000; Kaitany et al., 2001; Dorrance et al., 2003b). The effectiveness of \textit{Rps} genes has been estimated to last from eight to fifteen years (Schmitthenner, 1985). Thus, breeders cannot solely
rely on *Rps* genes due to their limited lifespans and the high pathotype diversity of *P. sojae* populations.

In contrast to *Rps* genes, quantitative resistance, or partial resistance, is controlled by multiple genes each contributing small effects to the trait (Kou and Wang, 2010). Partial resistance to *P. sojae* has been shown to be effective against all current races of *P. sojae* (Tooley and Grau, 1984; Schmitthenner, 1985). Unlike *Rps*-mediated resistance, partial resistance, as the name implies, is generally incomplete and allows some minimal pathogen growth. Soybean cultivars expressing only partial resistance without *Rps*-mediated resistance will have reduced symptoms from the pathogen, such as number of root infections, amount of colonization and reproduction, and an increased latent period (Kou and Wang, 2010). Multiple mechanisms for quantitative resistance have been previously hypothesized (Poland et al., 2009), sometimes involving plant architecture traits, such as root structure. Hence, partial resistance generally lacks the “gene-for-gene” interaction (Flor, 1955) of *Rps* gene mediated resistance and only moderate selection pressure is placed on the *P. sojae* population. Partial resistance is expected to be a more durable and stable resistance and has been shown to remain effective for several decades in other crops, such as several highly tolerant winter wheat cultivars that have been effectively controlling powdery mildew for over a half century (Shaner, 1973; Liu et al., 2001).

Improvement of partial resistance to *P. sojae* in N. American soybean cultivars can be achieved through the introgression of novel alleles contributing to partial resistance. Over 1,000 soybean plant introductions (PIs) were previously evaluated as potential novel sources of resistance to *P. sojae* and among those PIs originating from the Republic of Korea (S. Korea) were found to be enriched for high levels of partial resistance (Dorrance and Schmitthenner, 2000). High genetic diversity for *P. sojae* resistance may exist in plant introductions from S.
Korea as a result of co-evolution of soybean and *P. sojae* at this potential site of origin (Dorrance et al., 2007).

Identification of QTL against *P. sojae* resistance from these PIs have been limited to seven PIs, three originating from S. Korea, for which recombinant inbred lines have been generated (Nguyen et al., 2012; Lee et al., 2013a, 2013b, 2014). Two QTL were reported on chromosomes 11 and 13 in recombinant inbred lines (RILs) derived from a cross between S99-2281 (Susceptible) x PI 408105A (Resistant) (Nguyen et al., 2012). Recently, Lee et al. (2013a, 2013b, 2014) identified a total of 23 QTL. Nine QTL were identified on chromosomes 3, 4, 8, 10, 13, 15, and 18 in a PI 407861A(R) x OX20-8(S) population. In another RIL population using a different PI parent, PI 398841(R) x OX20-8(S), ten QTL were identified 1, 2, 3, 4, 7, 15, 18 and 20. Four novel QTL were additionally identified in a six-nested inbred population with OX20-8 (S) × PI 398841, PI 407861A, PI 427106, PI 427105B, PI 398297, and PI 417178 (R) on chromosomes 4, 9, 12, and 16, respectively. Given the success in identifying alleles contributing to partial resistance from Plant Introductions, it is pertinent to further evaluate a broader array of plant introductions for partial resistance and to identify the common alleles contributing to resistance.

Genome-wide association (GWA) mapping is the non-random association of alleles at different loci in a population. This method has several advantages over bi-parental mapping, including higher allele diversity and resolution within natural populations versus recombinant inbred lines (Xue et al., 2013; Holland 2007). However, there are several disadvantages to using GWA mapping including that it is less likely to identify rare alleles in the population and requires a high density of markers (Thornsberry et al., 2001; Yu et al., 2006; Mezmouk et al., 2011). Thus, expectations from GWA mapping are that common alleles in marker rich regions with large
effect are identified. In soybean, GWA mapping has been applied to populations of plant introductions to confirm or identify novel QTL for a number of quantitative traits, such as *Sclerotinia* stem rot resistance (Elmer et al., 2015), protein and oil (Hwang et al., 2014), and yield component traits (Hao et al., 2012).

The recent availability of over 50,000 SNP markers from the Soy50KSNP chip assayed on nearly 20,000 accessions from the soybean National Plant Germplasm System have made GWA mapping on large populations of plant introductions into tractable studies (Song et al., 2013). The use of the Soy50KSNP beadchip for GWA mapping has been validated by recent studies. While protein and oil content of soybean seed had been intensely studied by traditional QTL analysis in bi-parental crosses, the re-identification by GWA mapping of QTL for protein and oil content in PIs genotyped with SoySNP50K chip served to validate GWA analyses using this genotypic data set (Hwang et al., 2014). The objective of this study was to 1) evaluate 1,398 S. Korean PIs for partial resistance to *P. sojae*, 2) to identify QTL by GWA mapping, and 3) to compare these QTL to known loci controlling resistance to *P. sojae*.

**Materials and Methods**

*Seed Material*

A collection of 1,398 PIs from S. Korea were obtained from the National Plant Germplasm System, consisting of fifty seeds per inbred line ranging from maturity groups I to IV. In addition to these lines, checks (‘Conrad’, ‘Sloan’, ‘OX20-8’, ‘Williams79’, ‘Resnik’, ‘L83-570’, PI 398841, and PI 407861A) with known, varied levels of resistance to *P. sojae* were included in the experimental design. All seeds were surface sterilized following a chlorine gas protocol adopted from Olhoft et al. (2006) prior to disease assays.
**Disease Assays**

Pathogenicity tests were conducted prior to the phenotypic disease assay for selection of two \textit{P. sojae} based on a hypocotyl inoculation test (Dorrance et al. 2008) and tray test (Olah and Schmitthenner, 1985; McBlain et al., 1991) for high virulence and aggressiveness as described in Chapter 2. Isolates \textit{OH12108_6.3 (OH121)} (\textit{vir 1a, 1b, 1c, 1d, 1k, 2, 3a, 3c, 4, 5, 6, 7, and 8}) and \textit{C2S1 (vir 1a, 1b, 1d, 1k, 2, 3a, 3b, 3c, 4, 5, 6, 7, and 8}) were identified as virulent pathotypes for eliminating \textit{Rps}-mediated responses such that partial resistance in the layer test could be observed.

Phenotypic disease assays were conducted on 1,398 PIs originating from S. Korea following a layer test protocol adopted by Dorrance \textit{et al.} (2008) to evaluate for partial resistance to \textit{P. sojae} as described in Chapter 2. Plant height averaged from three seedlings per cup (PH), adjusted root weight (RW), adjusted shoot weight (SW), and root rot score (RRS) were measured two weeks after planting for the inoculated (I) and non-inoculated treatments (N). Root rot was rated on the 1-9 scale adopted from Dorrance \textit{et al.} (2008). Fresh weights of the roots and shoots were adjusted by dividing by the total number of plants per cup. The experiment was repeated twice in time for each isolate. The 1,398 PIs were first evaluated with \textit{C2S1}, the less aggressive of the two isolates. PIs which exhibited little no root rot (average IRRS < 1.5) were removed from the subsequent disease assays with \textit{OH121}. PIs with non-inoculated average root rot scores > 1.5 were removed from the dataset, eliminating possible disease due to contamination or seed-borne pathogens. A total of 804 PIs with an average IRRS $\geq$ 1.5 for both isolates and a mean NRRS < 1.5 were further analyzed for heritability and GWA mapping.
Phenotypic Data Analysis

Best linear unbiased predictors (BLUPS) were generated for each PI for IRRS, IRW, ISW, IPH, NRW, NSW, and NPH using PROC MIXED procedure in the software SAS (SAS 9.3, SAS Institute Inc., Cary, NC, USA) (Stroup, 1989) as described in Chapter 3. Broad-sense heritability ($H^2$) was calculated for each trait as follows: $H^2 = \sigma_g^2 / (\sigma_g^2 + \sigma_{gi}^2/i + \sigma_e^2/ir)$, where $\sigma_g^2$ represents the genetic variance, $\sigma_{gi}^2$ represents the genotype × isolate variance, $\sigma_e^2$ represents the error variance, $r$ is the number of experimental replicates, and $i$ is the number of isolates.

Genotypic analysis

Genotypic data collected using the Illumina Infinium SoySNP50K iSelect BeadChip developed by the Beltsville, MD, USDA Soybean Genomics and Improvement Lab was downloaded from Soybase (http://www.Soybase.org). The genotypic data consisted of 52,041 SNPs (Song et al., 2013). Monomorphic markers and markers with > 5% missing data or >10% missing plus heterozygous allele calls were removed. A total of 19,138 markers were polymorphic, including 11,126 with minor allele frequencies (MAF) ≥ 5%, and were included in the association analysis (Song et al., 2013). Missing genotypes were imputed with fastPHASE (Scheet and Stephens, 2006). Genome-wide association mapping was carried out on the S. Korean population as previously described in Chapter 3.

Association Analysis

Analyses of the phenotypic and genotypic data sets for the 804 non-immune cultivars were conducted as described in Chapter 3. Briefly, GWA mapping was carried out using the GAPIT (Genome Association Prediction Tool) package (Lipka et al., 2012) in R with a compressed
mixed linear model (CMLM) with population parameters previously determined and the number of principle components used as covariates determined by Bayesian Information Criterion (Zhang et al., 2010). The significance threshold for marker-trait associations was determined by a modified Bonferroni adjustment in which the effective number of independent tests \( M_{\text{eff}} \) was calculated using simpleM (Gao et al., 2008). Genome-wide threshold levels for each of the two datasets were determined by \( \alpha/M_{\text{eff}} \), where \( \alpha=0.05 \) (Miller, 1981).

Results

Phenotypic Data Analysis

Phenotypic data from 518 lines with no observable root rot (IRRS ≤ 1.5) were excluded from analysis in order to eliminate any possible Rps-mediated resistance response. The remaining 804 PIs had a mean RRS of 3.5 which ranged from 1.5 to 8. Based on comparison of IRRS BLUP values to the mean solution of fixed effects of IRRS for resistant checks (L83-570, Williams 79, PI 407861A, PI 398851, and Conrad), 29% of the PIs were more resistant than check cultivars. Fifty-two percent of PIs were more susceptible than the susceptible check cultivars (Sloan and OX 20-8). The distributions of IRW, ISW, and IPH were normal and IRRS was skewed to the left (Figure 4.1). All four inoculated traits were strongly correlated to each other, where IRW, ISW, and IPH were positively correlated, and IRRS, in which a lower value indicates greater resistance, had a negative correlation to IRW, ISW, and IPH (Figure 4.1). There was significant genetic variance for the four traits in both inoculated and non-inoculated treatments (Table 4.1). The heritability for the four inoculated traits ranged from moderately low at 0.334 for IRW to moderately high at 0.605 for ISW (Table 4.1). In an analysis of assays with each isolate separately, C2S1 had lower heritability (Table 4.1) and lower disease incidence compared to
OH121, where the mean of the raw IRRS for C2S1 was 2.85 and the mean of the raw IRRS for OH121 was 4.13. The non-inoculated traits were all highly heritable, greater than 0.720.

Association Analysis

A total of 19,138 of polymorphic markers were used to carry out GWA mapping. Due to linkage disequilibrium among markers, they could be condensed into 12,313 effectively distinct markers. Principle Component Analysis of genotypic data indicated that the population structure of the 804 non-immune lines revealed three non-distinct groups (Figure 4.2). We applied a stringent correction for multiple testing in which marker trait associations were considered significant when the p-value was less than 4.06 × 10⁻⁶ [⁻log(p-value) = 5.39], or 0.05 divided by the effective number of markers (Gao et al., 2008). A total of 16 significant marker-trait associations were identified for IRRS and IRW (Table 4.1). These located to three different genomic regions on chromosomes 3, 13, and 19 (Figure 4.3).

Genomic regions on chromosomes 3 and 19 were identified with significant marker-trait associations for IRRS. Five SNPs on chromosome 3 (ss715585728, ss715585712, ss715586321, ss715586320, ss715586376) were located within a 400,000 bp region (3.9-4.3 Mbp) (Figure 4.3A). ss715586376 was the most significantly associated SNP with a –log(p-value) of 8.49 (higher value indicates a stronger association with the trait) (Table 4.1). Ten SNPs on chromosome 19 (ss715635897, ss715635934, ss715636056, ss715636059, ss715636064, ss715636073, ss715636076, ss715636077, ss715636083, ss715636084) were significantly associated with IRRS and were located within a 1.6 Mbp region (49.4-50.7 Mbp) (Figure 4.3A), with all ten SNPs possessing a –log(p-value) of 5.74 (Table 4.1).
A single SNP (ss715614689) at 28.7 Mbp on chromosome 13 was significantly associated with IRW. The –log(p-value) of the association was 5.84 (Figure 4.3B, Table 4.1). While no significant marker-trait associations were identified for ISW or IPH (Figures 4.3B and 4.3C), for IPH we noted near significance of markers at the chromosome 13 QTL for IRW and at the chromosome 3 QTL for IRRS, with –log(p-values) of 5.32 and 4.95, respectively.

No significant-marker trait associations were found in the NRW or NPH. Five genomic regions were identified with significant marker associations for NSW on chromosomes 2 (3.4 Mb- 4.6 Mb), 3 (5.2Mb- 5.6 Mb), 4 (6.1 Mb- 6.6 Mb), 17 (8.5 Mb), and 18 (51.7 Mb- 53.0 Mb) (Figure 4.4, Table 4.2). None of the QTL regions for NSW were coincident with any of the QTL identified for the inoculated traits.

Co-localization of significant SNPs with QTL, Rps and candidate genes

The QTL for IRRS on chromosome 3 is coincident with Rps1, RpsYu25, and RpsUn1 and nearby the mapped position of Rps7 (Weng et al., 2001; Sugimoto et al., 2008; Sun et al., 2011; Lin et al., 2013) (Figure 4.5A). Additionally, it is coincident with a previously identified QTL for lesion length in a tray test of partial resistance to P. sojae with the resistance allele from a plant introduction originating from S. Korea (Lee et al., 2013a). Several QTL on chromosome 3 associated with resistance to the necrotrophic pathogen, Sclerotinia sclerotiorum (Arahana et al., 2001) coincide with the QTL identified in this study. A total of 25 genes are annotated within this genetic region (Glyma.Wm82.a2.v1). Four of these genes putatively encode a nucleotide binding (NB) domain characteristic of R-genes (McHale et al., 2006) (Figure 4.5A). In addition, there are two leucine-rich repeat-receptor like kinase (LRR-RLK)-encoding genes in this region,
which have been shown to play a role in basal defense and plant developmental responses (Gomez-Gomez and Boller, 2000; Zipfel et al., 2006; Kemmerling et al., 2011).

The QTL for IRW on chromosome 13 is coincident with Rps8 and with three previously identified QTL for partial resistance to P. sojae (Wang et al., 2010; Wang et al., 2012; Nguyen et al., 2012). Additionally, this QTL is coincident with a QTL for flooding tolerance (Nguyen et al., 2012) and a disease resistance QTL to Sclerotinia sclerotiorum (Arahana et al., 2001). A cluster of 14 genes which putatively encode NB domains are located from 7 to 10 Mb from the single significant marker defining this QTL (Figure 4.5B).

The significant markers for the IRRS QTL on chromosome 19 are located ~11 Mb from a previously identified QTL for partial resistance to P. sojae through a tray test disease assay (Wang et al., 2012) (Figure 4.5C). There no QTL associated with other diseases that are coincident with the QTL on chromosome 19; however there are several coinciding QTL for three morphological traits including leaflet length (Orf et al., 1999), leaflet shape and width (Kim et al., 2005), and development traits (Orf et al., 1999; Li et al., 2008; Yamanaka et al., 2001).

A total of 202 genes are within the genetic region of the QTL for IRRS on chromosome 19, 47 of which represent candidate genes for quantitative resistance, none with similarity to the canonical R-genes which encode NB or receptor-like proteins (Wang et al., 2008). Candidate genes within this region include two putative LRR-RLK-encoding genes, six putative chitinase-encoding genes, three genes putatively involved in flavonol biosynthesis (Hückelhoven, 2007), three genes putatively involved in the control of gene silencing (Waterhouse et al., 2001), a putative ribose 5-phosphate isomerase A-encoding gene, for which homologs in Arabidopsis have been shown to function in cellulose synthase (Howles et al., 2006) and accumulation was reduced in response to type-III secreted effectors from Pseudomonas syringae (Jones et al.,
2006), and a putative ammonium transporter for which homologs have been implicated in interactions with root endophytes (Helber et al., 2011; Lahrmann, 2013; Favre et al., 2014) and can act as a negative regulator of basal defense responses in *Arabidopsis* (Pastor et al., 2014). In addition, 16 genes are similar to genes which are regulated in response to salicylic acid, ethylene, jasmonic acid, abscisic acid, or auxin, six of which putatively encode transcription factors and three of which are involved in additional signal transduction (Bari and Jones, 2009; Kazan et al., 2009).

**Discussion**

*QTL identified in this study*

In this study, 1,398 plant introductions from S. Korea were evaluated for partial resistance to *P. sojae* via a layer test disease assay with two replications for each isolate. A total of 804 PIs were evaluated in GWA mapping, 518 lines were excluded due to lack of disease development, potentially due to *Rps* gene-mediate resistance. Few GWA mapping studies have been conducted on plant introductions for partial resistance to *P. sojae*. In the present study, 16 significant marker-trait associations were identified in three genomic regions located on chromosomes 3, 13, and 19. The number of markers and genetic regions identified are similar to those of other GWA mapping studies for disease resistance in soybean (Bao et al., 2014; Sun et al., 2014; Elmer et al., 2015). For example, in a smaller GWA study of partial resistance to *P. sojae*, Sun et al. (2014) identified four significant SSR markers (Satt301, Sat_222, Satt634-133, and Satt634-149) on chromosome 2 and 17, utilizing 175 accessions from China and 495 SSR markers, where Satt301 was located in a region with a previously identified QTL associated to *P. sojae* on chromosome 17 (Wang et al., 2010) and the regions on both chromosome 2 and 17...
were coincident with previously identified QTLs and \( R \)-genes for other pathogens. The co-localization QTL for partial resistance with other genes and QTL can inform hypotheses on mechanisms of partial resistance.

It has been posited that partial disease resistance can be controlled through developmental or morphological mechanisms, basal defense genes such as receptor like kinases (RLKs), production of antimicrobials, components of defense signal transduction pathways, or weak \( R \)-gene responses (Poland et al., 2009). QTL for partial resistance to \( P. \) sojae from this study and previous studies can be condensed into 18 genetic regions. Of the 18 QTL regions identified for partial resistance to \( P. \) sojae, four are coincident with \( R \)-genes, 16 are coincident with QTL for resistance to necrotrophic pathogens, one is coincident with QTL for resistance to biotrophic pathogens, and one is novel or co-localize only with other QTL for partial resistance to \( P. \) sojae (http://soybase.org; Burnham et al., 2003; Weng et al., 2007 Han et al., 2008; Li et al., 2010; Tucker et al., 2010; Wang et al., 2010; Wu et al., 2011; Wang et al., 2012; Nguyen et al., 2012; Lee et al., 2013a,b; Lee et al., 2014).

Inferences about mechanisms of partial resistance

While functional gene analysis is required to identify mechanisms, the co-localization of annotated genes and QTL can provide evidence in support of particular mechanisms. Mechanisms of partial resistance associated with developmental or morphological mechanisms are difficult to assess in this manner because there are not a limited number of clear defined pathways involved. However, the QTL identified on chromosomes 3, 13, and 19 are not coincident with any apparently relevant development or morphological QTL (http://soybase.org). Both the QTL on chromosome 3 and 19 are coincident with LRR-RLKs
characteristic of the pathogen associated molecular pattern (PAMP)-triggered immunity of basal defense as well as regulation of developmental processes (Gomez-Gomez and Boller, 2000; Zipfel et al., 2006) and thereby potentially adding support to the developmental/morphological and basal defense mechanistic hypotheses for partial resistance. The QTL on chromosome 19 is also associated with a suite of genes putatively involved in basal defense responses (Waterhouse et al., 2001; Jones et al., 2006; Pastor et al., 2014), signaling of defense response (Bari and Jones, 2009; Kazan et al., 2009), and production of antimicrobials (Hückelhoven, 2007), providing candidate genes to test the mechanism of resistance at this locus, but no clarity as to the precise mechanism.

In support of the weak R-gene hypothesis, the QTL identified on chromosomes 3 and 13 as well as a previously identified QTL on chromosome 3 (Lee et al., 2013a) are coincident with or nearby \textit{Rps} genes and putative NB-encoding (Lohnes and Schmitthenner, 1997; Diers et al., 1992; Sandhu et al., 2005). From the QTL identified on chromosome 19 as well as quantitative disease loci from other studies, it is clear that not all QTL for partial resistance can be weak forms of \textit{R}-genes. However, functional analysis of NB-encoding genes could provide mechanistic information for a subset of QTL for partial resistance.

While a number of candidate genes have been identified, further work should be done to determine linkage disequilibrium (LD) blocks associated with the significant markers. This is will allow a more thorough consideration of genes which may be outside of these regions, but in LD with the markers and potential conferring partial resistance.
**Limitations of P. sojae isolates**

Heritability of traits when only assays with the C2S1 were considered were much lower than with the full dataset or with the OH121 isolate alone. While the differences in heritability of each isolate is similar to previous reports (Nguyen et al., 2012) and can likely be attributed to the reduced disease development from the C2S1 isolate it emphasizes the need to carry out assays for partial resistance with multiple isolates. Although the C2S1 isolate was virulent in the hypocotyl test and moderately aggressive in a tray test, the aggressiveness of *P. sojae* isolates can fluctuate depending upon the phenotypic disease assay (Wang et al., 2012).

While 518 lines were immune to disease and therefore did not contribute to the GWA mapping in this study, they are potential sources of novel *Rps* genes. Additionally, on the basis of root rot score, root weight, shoot weight, and plant height, the most resistant plant introductions (Table 4.3) were identified as potential sources of partial resistance in breeding programs. Further haplotype analysis should be carried out in order to identify haplotypes associated with high levels of resistance at the three QTL identified here. Markers designed to these haplotypes can be used for MAS. Improved partial resistance will allow more durability against aggressive isolates of *P. sojae*.
**Tables and Figures**

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Trait†</th>
<th>$\sigma_g^2$</th>
<th>$\sigma_i^2$</th>
<th>$\sigma_{gi}^2$</th>
<th>$\sigma_k^2$</th>
<th>$\sigma_r^2$</th>
<th>$\sigma_e^2$</th>
<th>$H^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>C2S1 and OH121‡</strong></td>
<td>IRRS</td>
<td>0.29***</td>
<td>0.83</td>
<td>0.22***</td>
<td>0.14***</td>
<td>0.00</td>
<td>1.52***</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>IRW</td>
<td>0.01***</td>
<td>0.00</td>
<td>0.01**</td>
<td>0.01***</td>
<td>0.03</td>
<td>0.08***</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>ISW</td>
<td>0.05***</td>
<td>0.00</td>
<td>0.01**</td>
<td>0.01***</td>
<td>0.00</td>
<td>0.10***</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>IPH</td>
<td>0.65***</td>
<td>0.00</td>
<td>0.82***</td>
<td>0.40***</td>
<td>2.34</td>
<td>2.75***</td>
<td>0.71</td>
</tr>
<tr>
<td><strong>OH121</strong></td>
<td>IRRS</td>
<td>0.75***</td>
<td>NA§</td>
<td>NA</td>
<td>0.05*</td>
<td>0.00</td>
<td>1.51***</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>IRW</td>
<td>0.03***</td>
<td>NA</td>
<td>NA</td>
<td>0.00*</td>
<td>0.00</td>
<td>0.06***</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>ISW</td>
<td>0.07***</td>
<td>NA</td>
<td>NA</td>
<td>0.00*</td>
<td>0.00</td>
<td>0.07***</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>IPH</td>
<td>1.43***</td>
<td>NA</td>
<td>NA</td>
<td>0.47*</td>
<td>0.34</td>
<td>3.75***</td>
<td>0.43</td>
</tr>
<tr>
<td><strong>C2S1</strong></td>
<td>IRRS</td>
<td>0.24**</td>
<td>NA</td>
<td>NA</td>
<td>0.19**</td>
<td>0.00</td>
<td>1.55***</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>IRW</td>
<td>0.01**</td>
<td>NA</td>
<td>NA</td>
<td>0.01**</td>
<td>0.06</td>
<td>0.09***</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>ISW</td>
<td>0.05***</td>
<td>NA</td>
<td>NA</td>
<td>0.01**</td>
<td>0.01</td>
<td>0.12***</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>IPH</td>
<td>1.45***</td>
<td>NA</td>
<td>NA</td>
<td>0.39**</td>
<td>4.21</td>
<td>1.88***</td>
<td>0.61</td>
</tr>
<tr>
<td><strong>None</strong></td>
<td>NRW</td>
<td>0.04***</td>
<td>NA</td>
<td>NA</td>
<td>0.02***</td>
<td>0.06</td>
<td>0.06***</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>NSW</td>
<td>0.09***</td>
<td>NA</td>
<td>NA</td>
<td>0.01**</td>
<td>0.02</td>
<td>0.10***</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>NPH</td>
<td>1.15***</td>
<td>NA</td>
<td>NA</td>
<td>1.28***</td>
<td>4.89</td>
<td>1.90***</td>
<td>0.72</td>
</tr>
</tbody>
</table>

**Table 4.1.** Variance components and heritability ($H^2$) of each trait. Data was analyzed with both isolates together and for each isolate separately. Significance of each variance component are indicated with *, **, and *** for p-values less than 0.05, 0.01, and 0.001, respectively. Genetic variance ($\sigma_g^2$), variance of isolate ($\sigma_i^2$), genotype by isolate variance ($\sigma_{gi}^2$), rep within block variance ($\sigma_k^2$), variance of rep ($\sigma_r^2$), and residual variance ($\sigma_e^2$) components are outputs from the SAS model that was generated for extracting genotypic best linear unbiased predictors (BLUPs).


‡ **OH121: OH121086.3**

§ Not applicable
<table>
<thead>
<tr>
<th>Trait</th>
<th>Marker</th>
<th>Chr.</th>
<th>Chr. position (bp)</th>
<th>Minor Allele Frequency</th>
<th>R²</th>
<th>p-value</th>
<th>-log(p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inoculated Root Rot Score</td>
<td>ss715586376</td>
<td>3</td>
<td>4315512</td>
<td>0.2</td>
<td>0.04</td>
<td>3.27E-09</td>
<td>8.49</td>
</tr>
<tr>
<td></td>
<td>ss715585728</td>
<td>3</td>
<td>3865730</td>
<td>0.18</td>
<td>0.03</td>
<td>7.26E-08</td>
<td>7.14</td>
</tr>
<tr>
<td></td>
<td>ss715585712</td>
<td>3</td>
<td>3852888</td>
<td>0.18</td>
<td>0.03</td>
<td>9.14E-08</td>
<td>7.04</td>
</tr>
<tr>
<td></td>
<td>ss715586321</td>
<td>3</td>
<td>4277380</td>
<td>0.21</td>
<td>0.03</td>
<td>1.04E-07</td>
<td>6.98</td>
</tr>
<tr>
<td></td>
<td>ss715586320</td>
<td>3</td>
<td>4276534</td>
<td>0.21</td>
<td>0.03</td>
<td>1.77E-07</td>
<td>6.75</td>
</tr>
<tr>
<td>Inoculated Root Rot Score</td>
<td>ss715635897</td>
<td>19</td>
<td>49121258</td>
<td>0.01</td>
<td>0.03</td>
<td>1.82E-06</td>
<td>5.74</td>
</tr>
<tr>
<td></td>
<td>ss715635934</td>
<td>19</td>
<td>49461582</td>
<td>0.01</td>
<td>0.03</td>
<td>1.82E-06</td>
<td>5.74</td>
</tr>
<tr>
<td></td>
<td>ss715636056</td>
<td>19</td>
<td>50544363</td>
<td>0.01</td>
<td>0.03</td>
<td>1.82E-06</td>
<td>5.74</td>
</tr>
<tr>
<td></td>
<td>ss715636059</td>
<td>19</td>
<td>50555433</td>
<td>0.01</td>
<td>0.03</td>
<td>1.82E-06</td>
<td>5.74</td>
</tr>
<tr>
<td></td>
<td>ss715636064</td>
<td>19</td>
<td>50604933</td>
<td>0.01</td>
<td>0.03</td>
<td>1.82E-06</td>
<td>5.74</td>
</tr>
<tr>
<td></td>
<td>ss715636073</td>
<td>19</td>
<td>50663466</td>
<td>0.01</td>
<td>0.03</td>
<td>1.82E-06</td>
<td>5.74</td>
</tr>
<tr>
<td></td>
<td>ss715636076</td>
<td>19</td>
<td>50666563</td>
<td>0.01</td>
<td>0.03</td>
<td>1.82E-06</td>
<td>5.74</td>
</tr>
<tr>
<td></td>
<td>ss715636077</td>
<td>19</td>
<td>50668662</td>
<td>0.01</td>
<td>0.03</td>
<td>1.82E-06</td>
<td>5.74</td>
</tr>
<tr>
<td></td>
<td>ss715636083</td>
<td>19</td>
<td>50680414</td>
<td>0.01</td>
<td>0.03</td>
<td>1.82E-06</td>
<td>5.74</td>
</tr>
<tr>
<td></td>
<td>ss715636084</td>
<td>19</td>
<td>50681263</td>
<td>0.01</td>
<td>0.03</td>
<td>1.82E-06</td>
<td>5.74</td>
</tr>
<tr>
<td>Inoculated Root weight</td>
<td>ss715615031</td>
<td>13</td>
<td>30766058</td>
<td>0.42</td>
<td>0.02</td>
<td>1.44E-06</td>
<td>5.84</td>
</tr>
</tbody>
</table>

**Table 4.2.** Significant marker-trait associations with identified genomic regions and chromosome location for each inoculated trait and model.
<table>
<thead>
<tr>
<th>Plant introduction</th>
<th>Inoculated Root Rot Score</th>
<th>Inoculated Root Weight</th>
<th>Inoculated Shoot Weight</th>
<th>Inoculated Plant Height</th>
</tr>
</thead>
<tbody>
<tr>
<td>PI 399006</td>
<td>1</td>
<td>1.24</td>
<td>1.33</td>
<td>10.97</td>
</tr>
<tr>
<td>PI 408090</td>
<td>1</td>
<td>1.2</td>
<td>1.4</td>
<td>8.9</td>
</tr>
<tr>
<td>PI 398687</td>
<td>1</td>
<td>1.14</td>
<td>1.26</td>
<td>9.63</td>
</tr>
<tr>
<td>PI 398299</td>
<td>1</td>
<td>1.12</td>
<td>1.17</td>
<td>8.15</td>
</tr>
<tr>
<td>PI 424518</td>
<td>1</td>
<td>1.03</td>
<td>2.33</td>
<td>9.47</td>
</tr>
<tr>
<td>PI 408319A</td>
<td>1</td>
<td>1</td>
<td>1.18</td>
<td>11.48</td>
</tr>
<tr>
<td>PI 398922</td>
<td>1</td>
<td>0.95</td>
<td>0.85</td>
<td>8.05</td>
</tr>
<tr>
<td>PI 424555B</td>
<td>1</td>
<td>0.94</td>
<td>0.98</td>
<td>11.07</td>
</tr>
<tr>
<td>PI 339868F</td>
<td>1</td>
<td>0.9</td>
<td>1.02</td>
<td>10.75</td>
</tr>
<tr>
<td>PI 424169B</td>
<td>1</td>
<td>0.89</td>
<td>1.31</td>
<td>11.67</td>
</tr>
<tr>
<td>PI 340010</td>
<td>1</td>
<td>0.84</td>
<td>0.96</td>
<td>8.63</td>
</tr>
<tr>
<td>PI 399004</td>
<td>1</td>
<td>0.83</td>
<td>0.96</td>
<td>8.85</td>
</tr>
<tr>
<td>PI 399015</td>
<td>1</td>
<td>0.82</td>
<td>1.52</td>
<td>6.7</td>
</tr>
<tr>
<td>PI 408068A</td>
<td>1</td>
<td>0.81</td>
<td>1.11</td>
<td>9.52</td>
</tr>
<tr>
<td>PI 407981A</td>
<td>1</td>
<td>0.81</td>
<td>1</td>
<td>11.83</td>
</tr>
<tr>
<td>PI 408074A</td>
<td>1</td>
<td>0.8</td>
<td>1.2</td>
<td>7</td>
</tr>
<tr>
<td>PI 424277A</td>
<td>1</td>
<td>0.73</td>
<td>1.1</td>
<td>9.07</td>
</tr>
<tr>
<td>PI 398730</td>
<td>1</td>
<td>0.73</td>
<td>1.2</td>
<td>7.7</td>
</tr>
<tr>
<td>PI 398849</td>
<td>1</td>
<td>0.72</td>
<td>0.82</td>
<td>6.7</td>
</tr>
<tr>
<td>PI 398775</td>
<td>1</td>
<td>0.7</td>
<td>1.11</td>
<td>9.13</td>
</tr>
<tr>
<td>PI 398694</td>
<td>1</td>
<td>0.7</td>
<td>0.87</td>
<td>9.93</td>
</tr>
<tr>
<td>PI 407805D</td>
<td>1</td>
<td>0.69</td>
<td>0.61</td>
<td>8.5</td>
</tr>
<tr>
<td>PI 398371</td>
<td>1</td>
<td>0.69</td>
<td>0.72</td>
<td>8.2</td>
</tr>
<tr>
<td>PI 407806A</td>
<td>1</td>
<td>0.68</td>
<td>0.76</td>
<td>6.83</td>
</tr>
<tr>
<td>PI 398964</td>
<td>1</td>
<td>0.67</td>
<td>1.3</td>
<td>8.37</td>
</tr>
<tr>
<td>PI 407795B</td>
<td>1</td>
<td>0.66</td>
<td>0.75</td>
<td>10.17</td>
</tr>
<tr>
<td>PI 424342B</td>
<td>1</td>
<td>0.65</td>
<td>1.08</td>
<td>7.68</td>
</tr>
<tr>
<td>PI 398848</td>
<td>1</td>
<td>0.65</td>
<td>1.1</td>
<td>7.7</td>
</tr>
<tr>
<td>PI 398242</td>
<td>1</td>
<td>0.62</td>
<td>0.93</td>
<td>8.93</td>
</tr>
<tr>
<td>PI 408287</td>
<td>1</td>
<td>0.59</td>
<td>0.7</td>
<td>8.85</td>
</tr>
<tr>
<td>PI 398299</td>
<td>1</td>
<td>0.59</td>
<td>0.7</td>
<td>7.48</td>
</tr>
<tr>
<td>PI 398704</td>
<td>1</td>
<td>0.56</td>
<td>0.8</td>
<td>7.53</td>
</tr>
<tr>
<td>PI 424316</td>
<td>1</td>
<td>0.54</td>
<td>0.69</td>
<td>6.15</td>
</tr>
<tr>
<td>PI 398702</td>
<td>1</td>
<td>0.47</td>
<td>0.7</td>
<td>7.18</td>
</tr>
<tr>
<td>PI 423842</td>
<td>1</td>
<td>0.45</td>
<td>0.75</td>
<td>6.48</td>
</tr>
<tr>
<td>PI 424171B</td>
<td>1</td>
<td>0.45</td>
<td>0.62</td>
<td>6.33</td>
</tr>
<tr>
<td>PI 408272C</td>
<td>1</td>
<td>0.43</td>
<td>0.63</td>
<td>6.28</td>
</tr>
<tr>
<td>PI 408062</td>
<td>1</td>
<td>0.35</td>
<td>0.46</td>
<td>8.37</td>
</tr>
</tbody>
</table>

**Table 4.3.** South Korean plant introductions with the highest levels of resistance based on an IRRS of 1 for disease assays with both C2S1 and OH121086.3 isolates and the means of the other three inoculated traits.
Figure 4.1. Histogram (diagonal), scatterplot (lower), and Pearson’s correlation (upper) for the four inoculated traits. Numbers in histograms indicate checks (1 = Conrad, 2 = L83-570, 3=OX-20, 4=PI398841, 5=PI407861A, 6=Resnik, 7=Sloan, and 8=Will79).

† Check groups 2, 5 and 8 are not indicated for root rot score on the histogram.

‡ Check group 3 is not indicated for plant height on the histogram.

§ Check group 7 is not indicated on the histogram for all four traits.
Figure 4.2. Principal Component (PC) Analysis of genotypic data for 804 non-immune South Korean plant introductions.
Figure 4.3. Manhattan plot of the soybean genome depicting the extent of associations of 19,138 SNPs with inoculated root rot score (A), inoculated root weight (B), inoculated shoot weight (C), and inoculated plant height (D). The $-\log_{10}(p)$-value on the y-axis is a measure of the degree to which a SNP is associated with the trait. Chromosomes are displayed on the x-axis. The red horizontal line indicates the genome-wide threshold for significance.
Figure 4.4. Manhattan plot of the soybean genome depicting the extent of associations of 19,138 SNPs with non-inoculated root weight (A), non-inoculated shoot weight (B), and non-inoculated plant height (C). The $-\log_{10}(p\text{-value})$ on the y-axis is a measure of the degree to which a SNP is associated with the trait. Chromosomes are displayed on the x-axis. The red horizontal line indicates the genome-wide threshold for significance.
Figure 4.5. Manhattan plot of regions of the soybean genome with significant SNP associations with inoculated root rot score on chromosome 3 (A), inoculated root weight on chromosome 13 (B), and inoculated root rot score on chromosome 19 (C). The red horizontal line indicates the genome-wide threshold for significance. The genomic positions of candidate resistance genes encoding nucleotide-binding sites are highlighted in green. The genomic localizations of Rps genes are indicated by grey horizontal bars (Weng et al., 2001; Sugimoto et al., 2008; Sun et al., 2011; Lin et al., 2013; Demirbas et al., 2001; Sandu et al., 2005). The genomic localization of previously mapped QTL for partial resistance to *P. sojae* are indicated by black horizontal bars (Lee et al., 2013; Wang et al., 2012; Wang et al., 2010; Nguyen et al., 2012).
Bibliography


Allington WB, Chamberlain DW (1948) Brown stem rot of soybean. Phytopathology 38 793-802


Dorrance AE, McClure SA, deSilva A (2003b) Pathogenic diversity of *Phytophthora sojae* in Ohio soybean production fields. Plant Dis 87: 139-146


Flor HH (1955) Host parasite interaction in flax rust-its genetics and other implications. Phytopathology 45: 680-685


Harris, DL. 1964. Expected and predicted progress from index selection involving estimates of population parameters. Biometrics 20: 46–72


Hickey JM, Crossa J, Babu R, de los Campos G (2012) Factors affecting the accuracy of genotype imputation in populations from several maize breeding programs. Crop Sci 52: 654-663

Hildebrand AA (1948) Soybean diseases in Ontario. Soybean Digest 10: 16-17


Kumar S, Volz RK, Weskett R (2011) Genetic architecture of fruit quality traits in Malus x domestica (Borkh.) compared between own-rooted seedlings and vegetative propagules on ‘M. 9’ rootstock. Tree Genet Genomes 7: 1079-1088


Mamidi S, Lee RK, Goos JR, McClean PE (2014) Genome-wide association studies identifies seven major regions responsible for iron deficiency chlorosis in soybean (Glycine max). PloS one 9: e107469


Miller, RG (1981) Simultaneous statistical inference. 2nd ed. Springer Verlag, pp 6-8


Olah AF, Schmitthenner AF,( 1985) A growth chamber test for measuring Phytophthora root rot tolerance in soybean seedlings. Phytopathology 75: 546-548


St. Martin SK, Scott DR, Schmitthenner AF, McBlain BA (1994) Relationship between tolerance to Phytophthora rot and soybean yield. Plant Breed 113: 331-334


Thomison, PR, Kenworthy WJ, McIntosh MS (1990) Phomopsis seed decay in soybean isolines differing in stem termination, time of flowering, and maturity. Crop Sci 30: 183-188


Turner SD (2014) qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. biorXiv DOI: 10.1101/005165


Wells WC, Kofoid KD (1986) Selection indices to improve an intermating population of spring wheat. Crop Sci 26: 1104-1109


