EVALUATION OF SOYBEAN CYST NEMATODE (SCN) RESISTANCE IN PERENNIAL GLYCINE SPECIES AND GENOME-WIDE ASSOCIATION MAPPING AND GENOMIC PREDICTION STUDY FOR SCN RESISTANCE IN COMMON BEAN AND PREDICTION OF THE SHORT DISTANCE MOVEMENT OF SOYBEAN RUST UREDINIOSPORES THROUGH MACHINE LEARNING

BY

LIWEI WEN

DISSERTATION

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

Urbana, Illinois

Doctoral Committee:

Professor Glen L. Hartman, Chair and Director of Research
Associate Professor of Patrick J. Brown
Associate Professor of Kris N. Lambert
Assistant Professor of Leslie L. Domier

ABSTRACT

Since agriculture started, there have been numerous occasions when plant diseases of crops had severe impact on human activities. From the famine caused by potato late blight (*Phytophthora infestans*) in Ireland in 1846, to the dramatic economic loss caused by downy mildew of grapes (*Plasmopara viticola*) in the Mediterranean in 1865, to the loss of the valuable banana cultivar 'Gros Michel' caused by *Fusarium oxysporum* Schlect. f. sp. *cubense*, plant diseases have caused significant historical and economic importance. The goal of plant disease management is to reduce the economic and aesthetic damage caused by plant diseases, and the focus of my thesis centers around studying diseases and their pathogen in an effort to supplement long-term effective management strategies for important diseases of soybean.

Soybean cyst nematode (SCN; *Heterodera glycines;* HG) is a widely occurring and damaging pathogen with a wide host range. SCN is the leading cause of soybean yield loss in the US and it will likely become a major yield-limiting threat to common bean (*Phaseolus vulgaris* L.), another highly susceptible host of SCN. Developing resistant cultivars is the most cost-effective method for managing this disease. In the first chapter of my thesis, I focused on identifying additional sources of resistance to SCN in perennial *Glycine* species which can be potentially used for improving resistance of soybean to SCN. 13 perennial *Glycine* species of 282 PIs were inoculated with HG types 0, 2, and 1.2.3 first, and then 36 PIs out of this set were further evaluated by inoculating with HG type 1.2.3.4.5.6.7, a population that overcomes all the resistance genes in soybean. The *Glycine* species evaluated contains many PIs that are highly resistant to SCN with 10 species classified as immune or highly resistance to three HG types, indicating a much broader resistance in these PIs. With additional work on hybridizing the perennial *Glycine* species and soybean along with techniques of gene cloning and gene transfer,

many of the genes in the perennial *Glycine* species could be used to develop additional soybean genotypes with SCN resistance. In the second chapter of my thesis, genome-wide association study (GWAS) was used to detect SNPs significantly associated with SCN resistance in the core collection of *P. vulgaris* and to make genomic predictions (GPs) of SCN resistance to two HG types. GWAS identified SNPs that are significantly associated with resistance to two HG types, and GP for resistance to two SCN HG types achieved high prediction accuracy. The findings in this chapter demonstrated GWAS and GP as valuable tools for developing new resistant common bean varieties with SCN resistance in the future.

Epidemiology studies concerning the environmental and biological factors affecting disease entry, establishment and development are also extremely important for the successful management of diseases. The third chapter of my thesis focuses on developing mathematical models to predict the disease epidemic of soybean rust (*Phakopsora packyrhizi*), another devastating fungal disease of soybean with rapid establishment and development in the fields, using environmental and biological variables. Four machine learning models, including Absolute Shrinkage and Selection Operator (LASSO) method, zero-inflated Poisson/regular Poisson regression model, random forest, and neural network were built and compare to describe deposition of urediniospores collected in passive and active traps. The high prediction accuracy of some of the models demonstrated the applicability of machine learning in disease risk assessment, and the finding of this project is potentially helpful in guiding farmers to make proper and in-time disease management decisions.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor Dr. Glen Hartman for his careful guidance and continuous support during my Ph.D. study. I appreciate his patience with me in all circumstances and his encouragement for me when I was having difficulties in research and life. I am tremendously lucky to have Dr. Hartman as my mentor.

Besides my advisor, I would like to thank the rest of my thesis committee: Dr. Pat Brown, Dr. Les Domier, Dr. Kris Lambert and Dr. Darin Eastburn (Emeritus), for their insightful comments and suggestions, and warm encouragements. I specially want to thank Dr. Pat Brown for providing me tremendous help in my genotyping-by-sequencing experiment and with data analysis.

Thanks to Dr. Roger Bowen for his great help in developing the manuscript in the rust epidemiology manuscript. I appreciate his work and input in this project. My sincere gratitude also goes to my peer colleague Hao-Xun Chang for his effort and help in the GWAS study, and for his challenging questions which made the study deeper and more precise.

I also want to thank all my other lab mates, including senior colleagues Dr. James Haudenshield, Mr. Curt Hill, Dr. Doris Lagos-Kutz, Dr. Guirong Zhang, and Ms. Theresa Herman for their inspiring suggestions and comments during my study, and my fellow colleagues including Michelle Pawlowski, Jiaqi Guo, Chelsea Harbach, Hui-Ching Yang for the stimulating discussions, and making the lab friendly and welcoming. I also want to thank my friends Xing Wu, Wei Wei, Yiwen Xiang, Wei Zhong, Jiangmeng Zhang, Zhoutao Pei, Tinghao Guo, Yusheng Feng, for their support, encouragement, and love at all times.

Last but not the least, I want to thank my family, my mom, dad, and brother. I can't say enough how much I appreciate their love and support. I thank my parents, especially my mom, for supporting all my decisions in life and sacrificing themselves for my growth and self-development. I also want to thank my brother for taking care of and making company of my parents, I can't thank him enough for being such a great brother.

TABLE OF CONTENTS

CHAPTER 1. ACCESSIONS OF PERENNIAL *GLYCINE* SPECIES WITH RESISTANCE TO MULTIPLE TYPES OF SOYBEAN CYST NEMATODE (*HETERODERA GLYCINES*)[1]

ABSTRACT

Soybean cyst nematode (SCN; *Heterodera glycines;* HG) is a widely occurring and damaging pathogen of soybean that limits soybean production. Developing resistant cultivars is the most cost-effective method for managing this disease. Genes conferring SCN resistance in soybean have been identified; however, there are SCN populations that overcome known resistance genes. In order to identify additional sources of resistance and potentially new resistance genes, 223 plant introductions (PIs) of *G. tomentella* and 59 PIs of 12 other perennial *Glycine* species were inoculated with HG Types HG 0, HG 2, and HG 1.2.3, and then 36 PIs out of this set were further evaluated by inoculating with HG type 1.2.3.4.5.6.7, a population that overcomes all the resistance genes in soybean. Of 223 *G. tomentella* PIs evaluated, 86 were classified as resistant to three HG types, 69 as resistant to two HG types, and 22 as resistant to one HG type. Of the other 12 perennial *Glycine* species, all PIs of *G. argyrea* and *G. pescadrensis* were resistant to all three HG types. Of the 36 PIs challenged with HG type 1.2.3.4.5.6.7, 35 were resistant with 16 showing no cyst reproduction. Our study confirms that there are high levels of resistance to SCN among the perennial *Glycine* species. This represents an untapped resource for use in genetic studies and for improving resistance to SCN in soybean.

[1]This chapter was published on Plant Disease（ISSN: 0191-2917）and reprinted from Wen, L., Yuan, C., Herman, T.K., and Hartman, G.L. 2017. Accessions of perennial *glycine* species with resistance to multiple types of soybean cyst nematode (*Heterodera glycines*). Plant Dis. (ja). doi.org/10.1094/PDIS-10-16-1472-RE

INTRODUCTION

Soybean cyst nematode (*Heterodera glycines*) (SCN) is the most damaging pathogen to soybean [*Glycine max* (L.). Merr.] production, incurring about 2.2% losses in production averaged over Argentina, Brazil, and the USA, the top three soybean-producing countries in the world (Hartman 2015; Niblack and Riggs 2015). Soybean plants are damaged when young juvenile nematodes feed on the vascular system and complete their life cycle, producing eggs inside cysts that protrude from the roots (Niblack and Riggs 2015).

Developing soybean cultivars with resistance to SCN is one effective approach for managing SCN. Although a number of sources of resistance have been discovered in plant introductions (PIs) in the USDA Soybean Germplasm Collection, mainly one source, PI 88788, has been employed in breeding programs (Joos et al. 2013). Shifts in SCN populations have led to a decrease in resistance in soybean cultivars derived from PI 88788 (Colgrove et al. 2002; Niblack et al. 2003, 2008), increasing the need to utilize other sources of resistance. Other sources found in soybean, like PI 437654 and PI 89772 (Anand 1992; Arelli et al. 2015; Nickell et al. 1999), have not been fully effective due to the vast genetic diversity of SCN populations.

The genetic variation in SCN populations was initially described by a race scheme in which the separation of different SCN populations was based on comparing female reproduction on four resistant differential lines (Golden et al. 1970). This system also was used to classify soybean genotypes based on responses to the various races ranging from immunity, or no reproduction, to highly susceptible (Schmitt and Shannon 1992). However, with the extensive genetic diversity of *H. glycines* populations, the race scheme was shown to be inadequate for defining SCN populations. Therefore, the race scheme was replaced by a new scheme, which

uses seven soybean indicator lines to classify nematode populations to HG types (Niblack et al. 2002).

There are only a few studies that have reported searching plant species other than *G. max* for resistance to SCN, although the pathogen is known to reproduce on other legume plants (Riggs 1992). The genus *Glycine* is composed of two subgenera, *Glycine* and *Soja*. Soja includes *G. max* and *G. soja*, an annual plant closely related and cross-compatible with soybean. The subgenus *Glycine* includes 24 perennial *Glycine* species, which mostly originate from diverse geographical areas in Australia, and are not cross-compatible with soybean (Hymowitz and Hartman 2015). Riggs et al. (1998) evaluated a total of 8 perennial *Glycine* species with one accession of *G. argyrea, G. arenaria, G. canescens, G. curvata, G. cyrtoloba, G. latifolia,* and *G. microphylla,* and four accessions of *G. tabacina* and *G. tomentella* to three isolates of *H. glycines* and determined that all were resistant with most cases having no cyst production. The most extensive study evaluated 491 accessions of 12 perennial *Glycine* species for resistance to *H. glycines* HG type 0 and showed that all species, except *G. curvata* and *G. pindanica*, had at least one accession with an immune response (Bauer et al. 2007). In addition to SCN resistance, some perennial *Glycine* species have been reported with resistance to other soybean diseases including brown spot (Lim and Hymowitz 1987), Phytophthora root rot (Kenworthy 1989), powdery mildew (Mignucci and Chamberlain 1978), soybean rust (Hartman et al. 1992; Hymowitz 1995), Sclerotinia stem rot and sudden death syndrome (Hartman et al. 1999). Thus, there is a reservoir of potentially new resistance genes in the perennial *Glycine* species that needs additional research and exploitation. These sources of resistance may be especially useful in cases where the pathogen population, like that of SCN, is complex. To our knowledge, published reports of perennial *Glycine* species evaluated for SCN resistance have used either only one HG type to

characterize accessions (Bauer et al. 2007) or multiple HG types with a very limited number of accessions (Riggs et al. 1998). The objective of our study was to evaluate 13 perennial *Glycine* species against SCN HG 0, HG 2, HG 1.2.3, and HG 1.2.3.4.5.6.7. This information would provide a more comprehensive understanding of the resistance structure in perennial *Glycine* species to several populations of *H. glycines*.

MATERIALS AND METHODS

SCN source and HG type confirmation on a soybean differential set. SCN HG types 0, 2, and 1.2.3 were obtained from G. R. Noel, USDA-ARS, Urbana, IL and multiplied once on the roots of the susceptible soybean cultivar Lee 74 before beginning this study. The SCN HG Type 1.2.3.4.5.6.7 (TN23) was collected from a field in Missouri and maintained in the greenhouse by mass selection on PI 437654. The isolate was obtained from Dr. Colgrove at the University of Illinois, and had been multiplied on the resistant soybean accession PI 437654 just before use in our study.

The SCN HG types 0, 2, 1.2.3, and HG 1.2.3.4.5.6.7 were confirmed on the standard susceptible cultivar Lee 74 and seven soybean indicator lines, Peking, PI 88788, and PI 90763, PI 437654, PI 209332, PI 89772 and PI 548316 (Table 1.1; Niblack et al. 2002). For each HG type test, five seeds of each indicator line were planted in five polyvinyl tubes with steam-pasteurized Torpedo sand (100 cm^3), one seed per tube. The tubes were randomly arranged in plastic buckets (19 cm in diameter and 20 cm in height) filled with 5 L pasteurized Torpedo sand. A shallow hole was dug next to each one-week-old seedling to inoculate by pipetting 1ml of 2,000/ml egg solution of the assigned SCN HG type. Cysts were extracted and counted on each indicator line 35 days after inoculation and the female index (FI) was calculated based on

the formula: FI = (mean number of females on an indicator soybean line / mean number of females on standard susceptible) × 100. The HG type of each SCN population was determined according to the FI of each indicator line with FI < 10 indicating negative host compatibility and FI >10 indicating positive host compatibility. The resistance level was classified into four categories based on FI value: resistant with FI ≤ 10, moderately resistant with 10<FI ≤ 30, moderately susceptible with 30<FI ≤ 60, and susceptible with FI >60 (Schmitt and Shannon,1992). The plastic buckets with plants were kept in a water bath at 28 ºC in the greenhouse at 16 hr of fluorescent light per day during the experiment.

Seed source and seed preparation of perennial *Glycine* species. There were 223 *G. tomentella* PIs and 59 PIs of 12 other *Glycine* species obtained from the USDA ARS Soybean Germplasm Collection (www.ars-grin.gov) that were tested in three experiments. The first experiment consisted of 223 *G. tomentella* PIs. The second experiment consisted of 59 PIs of 12 *Glycine* species. The third experiment was based on a selection from the first two experiments. For all experiments, seeds of the perennial *Glycine* species were scarified (required for germination) by removing a tangential portion of the seed coat on the side opposite the hilum with a sharp razor blade. Each seed was placed in a rolled 6.5 mm diameter filter paper disk (No. 3 Whatman, Fisher Scientific) inside a single well of a 96-well ELISA plate (12 columns and 8 rows). Twenty µl of distilled water were added to each well, and the plates were covered and incubated at 25° C under high humidity with 16 hr of fluorescent light per day for 3 to 5 days until the seeds germinated.

Experimental set up and design. In the first and second experiment, PIs were inoculated with SCN HG 0, HG 2, and HG 1.2.3 (Table 1.2 and Table 1.3). In the third experiment, 36 PIs out of the 282 PIs tested were further evaluated by inoculating with HG type 1.2.3.4.5.6.7, a

population that overcomes all the resistance genes in soybean (Table 1.4). The soybean cultivar Lee 74 was included in these experiments as a susceptible check.

For the first and second experiment, each PI was inoculated with SCN HG 0, HG 2, and HG 1.2.3 separately in a completely randomized design with five replications for each treatment (PI entry and HG group). Seedlings germinated on filter paper were transplanted into clay pots (10 cm in diameter) filled with pasteurized Torpedo sand. A hole was dug between the edge of each pot the transplanted seedling and 15 cysts of similar size of each SCN HG type were placed into the hole and covered with sand at the time of sowing. Each pot was an experimental unit. The inoculated seedlings were grown in a greenhouse for 35 days at 27° ± 3 C. Supplemental illumination was provided by a mixture of high-pressure sodium vapor and metal halide lamps for a 16 h photoperiod. The five replications were completed individually over time.

For the third experiment, PIs were tested in water baths with controlled temperature using a randomized complete block design (RCBD) with 10 blocks (replications). Each block consisted 37 polyvinyl chloride tubes (100 cm^3) placed in two plastic buckets (19 cm dia. x 20 cm h) filled with 5 liters of pasteurized Torpedo sand (18 tubes in one bucket and 19 in the other). Seedlings of 36 perennial *Glycine* accessions and the standard susceptible check Lee 74 were randomly transplanted per block into pasteurized Torpedo sand, one plant per block. Each tube was an experimental unit with 10 replications for each accession. Seedlings were inoculated with 2,000 eggs three weeks after seedlings were transplanted. Two seedlings per entry were sampled at 5 days post inoculation (DPI), 10 DPI, and 15 DPI to evaluate nematode development on each perennial *Glycine* plant compared with the susceptible check Lee at each time point. Roots were washed, and stained with acid fuchsin (Byrd et al., 1983), and a dissecting microscope at 64X magnification was used to quantify juvenile nematodes of all stages (J2, J3, and J4) in the entire

root system. The remaining four plants of each PI were kept in the plastic buckets and harvested at 35 DPI for cyst counting. Plastic buckets with plants were set in a 28° C water bath in the greenhouse. Supplemental illumination was provided by a mixture of high-pressure sodium vapor and metal halide lamps for a 16 h photoperiod. The experiment was completed twice with a different randomization for the second run.

Cyst extraction and counts. Thirty-five days after inoculation, plants were cut at the soil line and the content of each pot or tube was poured into a plastic bucket, and about 2 liters of water was added to make a soil suspension. The soil suspension was mixed well with pressurized water by hand for about 10 seconds, and then sediments were allowed to settle for 5 seconds leaving cysts floating. The suspension was poured through an 850 μm sieve to remove larger debris, and then over a 250 μm sieve. Finally, retained cysts and accompanying sediments were washed into a clean 50 ml plastic tube.

Cysts collected in each 50-ml tube were decanted on to a glass Petri plate, marked with a 1 mm^2 grid, and individually counted under a dissecting microscope (Olympus SZX16, Tokyo, Japan) at 25X. The mean cyst counts were taken over four plants of each PI. Significant variation in root size among the tested 282 perennial PI accessions was observed in experiment 1. To adjust for this, root size at washing was recorded based on a visual scale from 1 to 10 with 1 = smallest to 10 = largest. The FI was adjusted to accommodate size differences in the entire root system using an adjusted female index (AFI) calculated with the formula: AFI = (mean number of females on tested accession / root size) / (mean number of females on standard susceptible/root size) × 100. Classification of resistance or susceptibility was as follows: resistant (R) = AFI 0-10; moderately resistant (MR) = AFI 11-30; moderately susceptible (MS) = AFI 31-60; and susceptible (S) = AFI 61 or greater. For the third experiment, FIs were calculated

and the data from the two runs were tested for homogeneity of error variances before combining data.

Data analysis. For the first and second experiment, the AFI was used to separate the PIs into response categories. In the third experiment, nematode (J2-J4) counts were log transformed after adding 0.5 to the raw count data. Analysis of variance was done to compare nematode (J2-J4) counts among different PIs and at different time points (5 DPI, 10 DPI, 15 DPI) with the aov function in R 3.2.1 (R Core Team, Vienna, Austria), and multiple comparison was done between each PI pairs and time point pairs using Tukey's HSD function in R 3.2.1. For cyst number comparison at 35 DPI in the third experiment, FI was used to separate the PIs into response categories.

RESULTS

SCN genotype confirmation. All SCN populations used in this study were confirmed using seven indicator lines (HG type test) (Table 1.1). The first population had FI > 10 on only PI 88788 and was classified as HG 2. The second SCN population had FI < 10 on all the seven indicator lines and was classified as HG 0. The third population had FI > 10 on Peking, PI 88788, and PI 90763 and was classified as HG 1.2.3. The fourth population had FI > 10 on all the seven indicator lines and was classified as HG 1.2.3.4.5.6.7.

Evaluation of perennial *Glycine* species for resistance to HG 0, HG 2, and HG 1.2.3. Among the 223 *G. tomentella* PIs evaluated, resistance to the three SCN types was identified in 86 PIs (Table 1.2), while only 6 PIs were susceptible to all three SCN types. One hundred and thirty-one PIs showed a differential reaction among the three SCN types from resistant to susceptible.

Of the 12 other perennial *Glycine* species evaluated, SCN reproduction was detected on accessions in 10 species, including first time reports for *G. arenaria* and *G. rubiginosa* as hosts to SCN. All PIs of *G. pescadrensis* and *G. argyrea* had no cyst reproduction regardless of the HG type. Of the 59 PIs evaluated, only 4 PIs (*G. microphylla*: PI 509487 and PI 509489; *G. tabacina*: PI 446974 and PI 509495) were susceptible to all three SCN HG types; all others were resistant to at least one SCN HG type (Table 1.3). Each of the 12 species had individual accessions that were resistant to all the three HG types of SCN, and 16 PIs in total that were resistant to all three SCN HG types.

Evaluation of selected *Glycine* species for resistance to HG 1.2.3.4.5.6.7. Based on germination rate, only 16 perennial *Glycine* lines were sampled and stained. Two seedlings of each PI were sampled at each time point, and juveniles of all stages (J2-J4) were counted in the entire root system for each PI**.** Experiment three was conducted twice, and the root staining data was combined for the two experiments. Nematode (J2-J4) counts in perennial species were compared among different PIs at different time points. Root staining of the 16 perennial *Glycine* lines and susceptible check Lee showed that at juvenile counts were significantly different among different PIs (P = 5.683e-13, Table 1.5). Lee 74 had significantly more cysts than all the perennial *Glycine* species, and among the perennial species, PI 505286, PI509451, PI 505235 had significantly more cysts than PI 509462, PI 573064, PI 339657, and PI 339655 (Table 1.6). At 5 DPI, all accessions harbored J2 (Fig. 1.1A), and the development status of nematode was not different between the susceptible check Lee74 and perennial species. At 10 DPI, both J2 and J3 stages were observed in the roots of Lee 74 with nearly 50% being J3s (Fig. 1.1B); however, only J2s were observed in the roots of perennial *Glycine* accessions, indicating slower development of SCN in perennial *Glycine* species than in Lee 74. At 15 DPI, when most of the

nematodes in the roots of Lee 74 had developed to J4s, most of the nematodes in the perennial *Glycine* accessions were at the J2 stage (Fig. 1.1C). Nematode number in the perennial species differed ($P < 0.01$) at different time points as counts (indiscriminative J2 to J4 stages) changed from 2.3 to 7.0 to 1.8 at 5, 10, and 15 DPI, respectively, with the 10 DPI count being different from 5 and 15 DPI. At 35 DPI, FI calculations on all 36 PIs except one were classified as resistant (Table 1.4). The one susceptible entry was *G. tabacina* PI 373990. Of the resistant PIs, 14 had no cyst production while 21 had low cyst counts (Table 1.4).

DISCUSSION

We evaluated 223 PIs of *G. tomentella* and 102 PIs of 12 other perennial species for resistance to SCN. In general, most of the PIs were classified as resistant to SCN and only a few as susceptible. These results are consistent with previous studies, in which immunity and high levels of resistance to SCN were identified in many PIs of the perennial *Glycine* species (Bauer et al. 2007; Riggs et al. 1998). In one of the earliest studies, Riggs et al. (1998) inoculated one PI from each of eight perennial *Glycine* species with four HG types (races 1,3,5 and 14) and found all PIs as immune or resistant except for PI 440956, which was moderately resistant to one HG type. In our study, 282 accessions from 13 perennial *Glycine* species were tested for resistance to three different SCN populations with PIs in each of the eight species used by Riggs et al. (1998), identified as resistant. Bauer et al. (2007) reported that 56% of the PI entries representing 12 perennial *Glycine* species were immune or resistant to one HG type. In our study, 10 of the 12 perennial *Glycine* species were classified as immune or highly resistant to three HG types, indicating a much broader resistance in these PIs.

It is clear that the germplasm of the perennial *Glycine* species contains many PIs that are highly resistant to SCN whereas the majority of the soybean germplasm is susceptible to SCN (Wang et al. 2003; Young 1990). In soybean, only a few PIs have been reported to be resistant to a specific or several SCN populations by several genes including *Rhg1-3*, *Rhg4*, and *Rhg5* (Concibido et al. 2004; Jiao et al. 2015). SCN populations are known to overcome the resistance source (Mitchum et al. 2007; Niblack et al. 2008). Although an immune response was identified in PI 437654 to all HG types tested (Anand et al. 1988), field populations isolated from Brazil were able to reproduce on Hartwig (derived from PI 437654) (Dias et al. 1998), and the isolate TN23 (later classified as HG 1.2.3.4.5.6.7) isolated from a field isolate in Missouri in 2005 reproduced on PI 437654 (Bekal et al. 2003). The high level and broad resistance to SCN in the perennial *Glycine* species suggests that they may be very valuable sources to broaden the genetic base of soybean and develop new cultivars with more durable resistance to SCN.

The differential reaction of PIs within some of the perennial *Glycine* species is interesting because these perennial *Glycine* species evolved primarily in Australia and presumably without SCN in many different niches (Doyle et al. 2004; Sherman-Broyles et al. 2014). Because of their susceptibility to SCN, the species are hosts with some remnant susceptibility genes from the common ancestor or ancestors of soybean and the perennial *Glycine* species. Since non-host resistance to a pathogen has been defined as an entire species being resistant to all genetic variation of the pathogen (Niks and Marcel 2009), the only species that fit this criterion in our study was G. *argyrea*. However, no conclusions can be made until more accessions of the species are evaluated. *Glycine arenaria* and *G. rubiginosa* were identified as hosts of SCN, which were not reported before. The extent of SCN hosts are not known, but other hosts outside the legume family have been reported (Riggs 1992; Venkatesh et al. 2000).

One of the resistance mechanisms used by non-host plants is to prevent penetration of the J2 stage worms into roots (Colgrove and Niblack 2008). In our study, we did not see evidence of this in the perennial *Glycine* species, but root staining showed that the number of nematode entering roots of the perennial *Glycine* species were fewer than that of Lee 74, and that the development of life stages past J2 was delayed or in the case of immunity did not occur.

Information on the genomics and the molecular characterization of SCN resistance in the perennial *Glycine* species is limited. A comparative genomics study showed that the linkage group (LG) 18 of *G. latifolia* had a high degree of collinearity with the chromosome 18 in soybean with no interchromosal rearrangements (Chang et al. 2014). A genome wide association mapping study identified one locus containing four SNPs located in the linked regions (5.8–12.0 Mb) of known *Rhg1* on chromosome 18 in *G. soja* (Zhang et al. 2016). These studies suggest that the resistance genes to SCN in soybean and its relatives may be conserved. It is well known that the genetic base of soybean is very narrow with an estimation of 50% genetic diversity and 81% less minor alleles lost in the process of domestication (Hyten et al. 2006; Zhou et al. 2015). Since a novel locus on chromosome 19 of *G. soja* was found associated with SCN resistance (Zhang et al. 2016), the perennial *Glycine* species may harbor novel SCN resistance genes not found in soybean, and thus would be important in developing novel SCN resistance in soybean. Despite the challenges in utilizing the genetic resources in perennial *Glycine* species due to genetic barriers, additional work on hybridizing the perennial *Glycine* species and soybean along with techniques of gene cloning and gene transfer may provide an avenue to incorporate genes from the perennial *Glycine* species into soybean. In an earlier study, based on an interspecific cross between soybean and *G. tomentella*, hybrid backcrossed lines were shown to be resistant to *H. glycines* (Riggs and Schmitt 1988). With so much resistance in the

perennial Glycine species, over 100 highly resistant PIs identified in our study, there is a good chance that some of the SCN resistant genes from these sources will be used in future breeding programs.

ACKNOWLEDGEMENTS

This study was supported by USDA Agriculture Research Service. The authors thank Dr. Colgrove and Dr. Noel (USDA-ARS) at the University of Illinois for providing nematodes for this study and Bob Campbell for conducting preliminary evaluations of many perennial *Glycine* species.

TABLES

Table 1.1. *Heterodera glycines* (HG) types confirmed on seven soybean indicator lines based on positive or negative host compatibility.

Indicator lines	Female index of each HG type[z]			
	HG 0	**HG 2**	**HG 1.2.3**	**TN 23**
PI 54840 (Peking)	-	-	+	+
PI 88788	-	+	+	+
PI 90763	-	-	+	+
PI 437654	-	-	-	+
PI 209332	-	-	-	+
PI 89772	-	-	-	+
PI 548316 (Cloud)	-	-	-	+
HG type determination	HG 0	HG 2	HG 1.2.3	HG 1.2.3.4.5.6.7

[z] "+" indicates positive host compatibility as the number of females produced was equal to or greater than 10% of the number produced on the standard susceptible cultivar Lee; "-" indicates negative host compatibility as the number of females is less than 10% of cultivar Lee. Ratings were based on five replications for each soybean line and each HG type.

Table 1.2. *Glycine tomentella* accessions evaluated for their response to three *Heterodera glycines* (HG) types

	Response[a]		
Plant introductions	**HG 0**	**HG 2**	**HG 1.2.3**
320548, 339655, 339657, 339663, 373987, 373988, 440999, 441001, 441004, 441006-441010, 446950, 446980, 446989, 446996, 483218, 483223, 483225, 483227, 483228, 499919, 499920, 499927, 499928, 499933, 499934, 499939, 499941, 505202, 505212, 505224, 505225, 505228, 505230, 505231, 505233, 505235, 505239, 505240, 505243, 505245, 505247, 505248, 505250, 505252, 505254, 505255, 505257, 505261, 505262, 505266-505268, 505275, 505276, 505281, 505286, 505292, 505297, 509491, 509501, 509502, 537294, 546957, 546958, 563881, 563892, 563893, 563896, 563898, 563899, 573056, 573058, 573064, 573068, 573070, 573071, 583969, 583971, 583975, 591606, 599422, 604474, 604487, 604488, 604491, 604494	R	R	R
393556, 393557, 446963, 446984, 446985, 483220, 483224, 483226, 499938, 499951, 505209, 505210, 505218, 505219, 505234, 505277, 505279, 505285, 509499, 563878	R	R	MR
499930, 505258, 505260, 583973	R	MR	R
373980, 399479, 440998, 446946, 446951, 446954, 446956, 446957, 446958, 446959, 446981, 446988, 446998, 483219, 499916, 499935, 499936, 505203, 505229, 505238, 505244, 505249, 505269, 505269, 563897, 573062, 573066, 573067, 591609, 604476-604479, 604484, 604489, 604490, 604492, 604493	MR	R	R
441011, 505213, 505206	R	MR	MR
441005, 446955, 446994, 499937, 499945, 499947, 505215, 505220, 505226, 505280, 505294, 583974, 604481	MR	R	MR
446991, 505264, 505298, 505304, 563895, 563976, 595823	MR	MR	R
441012, 446995, 505217, 505232, 505236, 505237, 505256, 505263, 505272, 505282, 505303, 509500	MR	MR	MR
233051, 320547, 37399, 505246, 505271	R	R	S
505221, 505242, 591605, and 595824	R	S	R
499990	R	MR	S
446960[b], 446992, 505205	MR	R	S
446962[c], 505216[c]	MR	S	R
505201[b]	S	R	MR
505222[c]	R	S	S
441000[d], 446982[d], 446993[bcd]	S	R	S
441003[cd], 446952[cd], 505214[cd], 563900[d]	S	S	R
446953[d], 505211[d], 505241[d], 505278[d]	S	MR	MR
446961, 446949[c], 563877	MR	S	MR
441002[cd], 505300, 505301[cd]	S	S	MR
446947[c], 573063[c]	MR	S	S
446983, 446948[bcd], 505274[bcd], 573059, 591607[b], 595822	S	S	S

[w] Resistant (R) = adjusted female index (AFI) 0-10; moderately resistant (MR) = AFI 11-30; moderately susceptible (MS) = AFI 31-60; and susceptible (S) = AFI 61 or greater. Rating based on five replications over time.
[x] Variable result in repeat for HG 1.2.3 from resistant to susceptible; classified as susceptible.
[y] Variable result in repeat for HG 2.5.7 from resistant to susceptible; classified as susceptible.
[z] Variable result in repeat for HG 0 from resistant to susceptible; classified as susceptible.

Table 1.3. *Glycine* species accessions evaluated for their response to three *Heterodera glycines* (HG) types[z]

		Response[a]		
***Glycine* spp.**	**Plant introduction**	**HG 0**	**HG 2**	**HG 1.2.3**
G. arenaria	505204	R	R	R
G. argyrea	509451, 505151	R	R	R
G. canescens	440942	MR	R	R
G. canescens	446934	MS	MS	R
G. canescens	509458	MS	S	MR
G. canescens	573046	MS	S	R
G. canescens	509457	R	R	R
G. canescens	440932	R	MS	R
G. canescens	509455	R	R	MS
G. canescens	505154	R	R	S
G. canescens	509456	R	R	MR
G. canescens	573044	R	MS	MS
G. canescens	509454	R	MR	MR
G. canescens	573045	S	S	MR
G. clandestina	440960	MR	S	R
G. clandestina	509461, 505163	MS	R	R
G. clandestina	509468	R	R	R
G. clandestina	509462, 509466	R	MS	R
G. curvata	505167	MS	R	R
G. cyrtoloba	440962	MS	R	R
G. cyrtoloba	373993, 509472	R	R	R
G. falcata	509473	MS	R	R
G. falcata	509475	R	MS	R
G. falcata	440975	S	S	R
G. latifolia	253238	MR	MS	R
G. latifolia	505181	MS	MR	R
G. latifolia	440978	R	R	R
G. latifolia	509478, 509476	R	MS	R
G. latifolia	509479	S	R	R
G. latifolia	509480	S	MS	R
G. microphylla	505188, 509486	R	R	R
G. microphylla	446939	R	MS	R
G. microphylla	505196	R	MR	R
G. microphylla	509485	R	S	R
G. microphylla	509489	S	S	MS
G. microphylla	509483	S	R	R
G. microphylla	509487	S	MS	MS
G. pescadrensis	505199	R	R	R
G. pescadrensis	505195, 505197	R	R	R
G. rubiginosa	440954	R	R	R
G. tabacina	509494	MR	MS	MR
G. tabacina	339661, 509493	MS	MS	R
G. tabacina	373986	MS	S	MR
G. tabacina	509496	R	R	R
G. tabacina	373990	R	R	R
G. tabacina	509497	R	R	MS
G. tabacina	440990	R	MS	MS
G. tabacina	509492	R	MS	MR
G. tabacina	509490	R	S	MR
G. tabacina	446974	S	MS	MS
G. tabacina	509495	S	MS	MS
G. tabacina	509498	S	MR	S

[z] Resistant (R) = (adjusted female index) AFI 0-10; moderately resistant (MR) = AFI 11-30; moderately susceptible (MS) = AFI 31-60; and susceptible (S) = AFI 61 or greater. Rating based on five replications over time.

Table 1.4. Number of cysts recovered from selected *Glycine* species plant introductions after inoculation with soybean cyst nematode HG 1.2.3.4.5.6.7

***Glycine* spp.**	**Plant introduction**	**Chromosome (2n)**	**Cyst count** [y]
G. argyrea	509451	40	0
G. canescens	440932	40	0
G. clandestina	509462	I [z]	0
G. clandestina	509468	40	0
G. latifolia	505181	I	1
G. microphylla	505188	40	1
G. falcata	509473	I	1
G. cyrtoloba	509472	40	2
G. latifolia	509479	40	2
G. microphylla	509486	I	3
G. canescens	573047	I	5
G. pescadrensis	505197	80	6
G. tabacina	446974	80	17
G. tabacina	373990	40	148
G. tomentella	339655	80	0
G. tomentella	339657	78	0
G. tomentella	441010	78,80	0
G. tomentella	446996	78	0
G. tomentella	483227	80	0
G. tomentella	505235	78	0
G. tomentella	505286	78	0
G. tomentella	509502	78 7	0
G. tomentella	573064	38	0
G. tomentella	604474	I	0
G. tomentella	320548	78,80	0
G. tomentella	505240	80	0
G. tomentella	583969	40	1
G. tomentella	604494	I	1
G. tomentella	499920	I	2
G. tomentella	573059	I	3
G. tomentella	446983	40	4
G. tomentella	563899	40	8
G. tomentella	546957	80	9
G. tomentella	446950	80	14
G. tomentella	373987	38	16
G. tomentella	505292	40	35
G. max	Lee 74	40	676
G. max	Hartwig	40	254
G. max	437654	40	224

[y] Numbers averaged over two runs with three replications per run.
[z] Chromosome number not known.

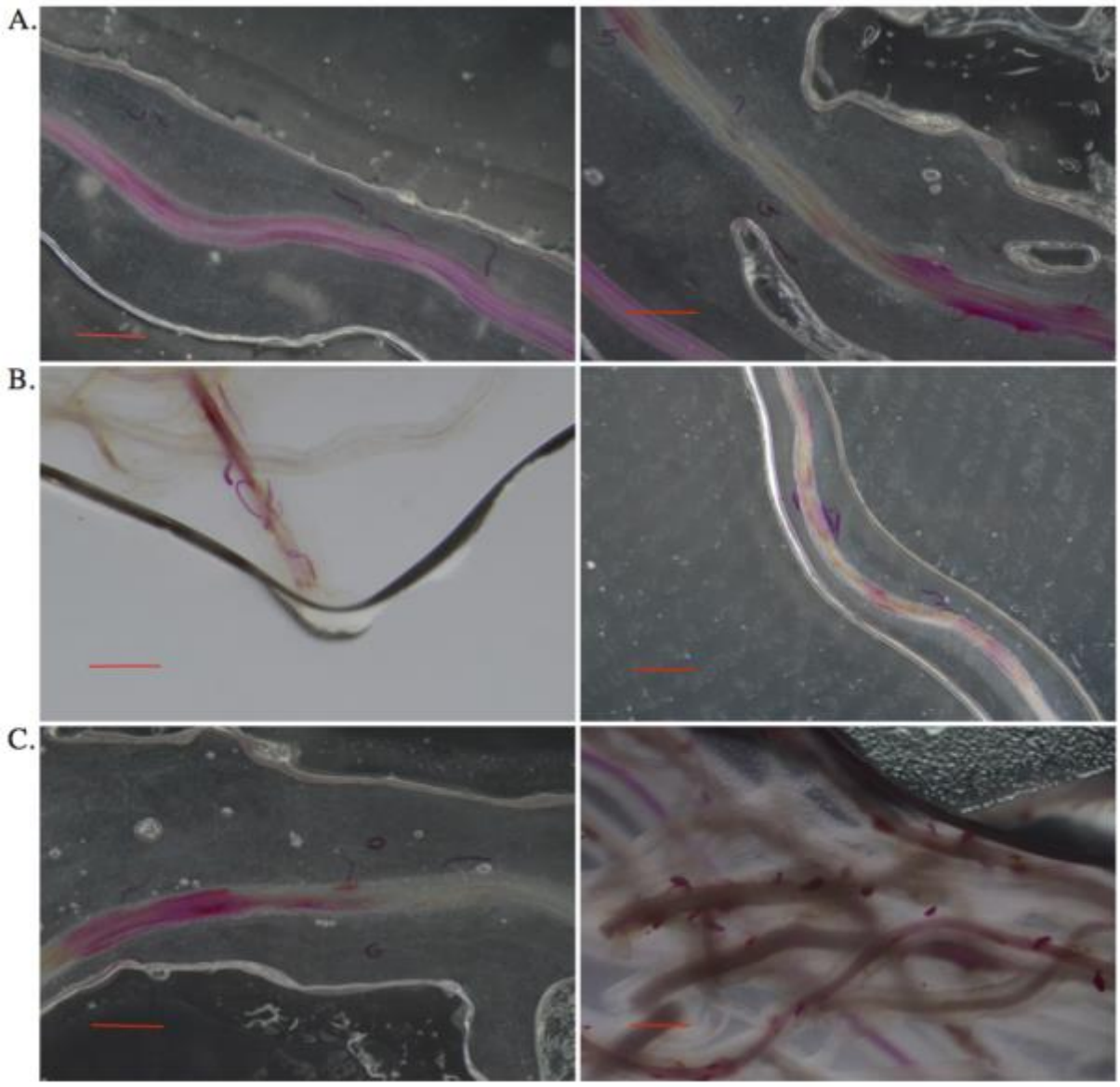
Table 1.5. Counts of *Heterodera glycines* juveniles in stained roots 5, 10, and 15 days after inoculation (DAI) and cysts counts 30 DAI on soybean cultivar Lee74 and on perennial *Glycine* species.

	Juvenile count (DAI)				
Entry[y]	**5**	**10**	**15**	**Cyst counts[z]**	
Lee74	61.0	88.0	134.0	89.1	a
PI 505286	0.5	13.0	2.0	4.9	b
Pi 509451	1.5	8.5	2.0	4.2	b
PI 505235	5.5	22.5	3.0	3.6	b
PI 441010	4.5	8.5	1.0	2.5	bc
PI 604474	1.5	6.0	3.0	2.5	bc
PI 505188	4.0	10.5	3.0	2.1	b-d
PI 505181	1.5	21.0	1.5	1.8	b-e
PI 440932	7.5	4.0	0.0	1.8	b-e
PI 509502	0.5	7.0	2.5	1.7	b-f
PI 446996	3.5	10.5	0.0	1.6	b-g
PI 509468	1.0	8.0	0.0	1.6	b-g
PI 509473	1.3	6.8	0.8	1.4	b-g
PI 483227	0.5	8.5	0.0	0.8	c-g
PI 509462	4.5	6.5	0.0	0.4	d-g
PI 573064	0.5	4.0	1.0	0.3	e-g
PI 339657	0.5	2.5	0.0	0.2	fg
PI 339655	0.5	0.5	0.0	0.2	g

[y]PI = plant introductions (PI); most are *G. tomentella* except for 509451 (G. *argyrea*), 509462 and 509468 (*G. clandestina*), 505181 (*G. latifolia*), and 505188 (*G. microphylla*).
[z]Mean nematode counts were back transformed for presentation; means with different letters are significantly different at $P = 0.05$ based on Tukey's HSD test. Each mean is based on 20 experimental units (two samples per 10 replications).

FIGURE

Fig. 1.1. Comparison of *Heterodera glycines* type 1.2.3.4.5.6.7 SCN development in the roots of perennial *Glycine* species and standard susceptible check Lee 74 stained with acid fuchsin. Fewer nematodes progressed from J2 to J3 or J4 in the perennial *Glycine* accessions. (A) A representative perennial *Glycine* accession, *G. tabacina* PI 505235 (left) and susceptible check Lee 74 (right) at 5 days post infection (DPI). (B) SCN development at 10 DPI in *G. tabacina* PI 446974 (left) and Lee 74 (right). (C) SCN development at 15 DPI in PI 505235 (left) and Lee 74 (right). Bar represents 500 microns in each image.



REFERENCES

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

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

Arelli, P. R., Pantalone, V., R., Allen, F. L., Mengistu, A., Fritz, L. A. 2015. Registration of JTN-5203 soybean germplasm with resistance to multiple cyst nematode populations. J. Plant Regist. 9:108-114.

Bauer, S., Hymowitz, T., and Noel, G. R. 2007. Soybean cyst nematode resistance derived from *Glycine tomentella* in amphiploid (*G. max* X *G. tomentella*) hybrid lines. Nematropica. 37:277-285.

Bekal, S., Niblack, T. L. and Lambert, K. N. 2003. A chorismate mutase from the soybean cyst nematode *Heterodera glycines* shows polymorphisms that correlate with virulence. Mol. Plant Microbe Interact. 16:439-446.

Byrd, D.W., Kirkpatrick, Jr.T., and Barker, K.R. 1983. An improved technique for clearing and staining plant tissues for detection of namatodes. J. Nematol. 15: 142-143.

Chang, H.-X., Brown, P., Lipka, A., Domier, L. L., and Hartman, G. L. 2016. Genome-wide association and genomic prediction identifies associated loci and predicts the sensitivity of *Tobacco ringspot virus* in soybean plant introductions. BMC Genomics 17:153:DOI 10.1186/s12864-12016-12487-12867.

Chang, S., Thurber, C. S., Brown, P. J., Hartman, G. L., Lambert, K. N. and Domier, L. L. 2014. Comparative mapping of the wild perennial *Glycine latifolia* and soybean (*G. max*) reveals extensive chromosome rearrangements in the genus *Glycine*. PloS One 9:e99427.

Colgrove, A. L., and Niblack, T. L. 2008. Correlation of female indices from virulence assays on inbred lines and field populations of *Heterodera glycines*. J. Nematol. 40: 39–45.

Colgrove, A. L., Smith, G. S., Wrather, J. A., Heinz, R. D., and Niblack, T. L. 2002. Lack of predictable race shift in *Heterodera glycines*- infested field pots. Plant Dis. 86:1101–1108.

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

Dias, W. P., Silva, J. F. V., Kiihl, R. A. D., Hiromoto, D. M., and Abdelnoor, R. V. 1998. Quebra da resistência da cv. Hartwig por população de campo do nematóide de cisto da soja (*Heterodera glycines*). Pesq. Agropec. Bras. 33:971-974.

Doyle, J. J., Doyle, J. L., Rauscher, J. T., and Brown, A. H. D. 2004. Evolution of the perennial soybean polyploid complex (*Glycine* subgenus *Glycine*): a study of contrasts. Bio. J. Linnean Soc. 82: 583-597.

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

Hartman, G. L. 2015. Worldwide importance of soybean pathogens and pests. Pages 4-5 in: Compendium of Soybean Diseases and Pests, G. L. Hartman, J. C. Rupe, E. F. Sikora, L. L. Domier, J. A. Davis and K. L. Steffey, eds. American Phytopathological Society, St. Paul.

Hartman, G. L., Gardner, M. E., Hymowitz, T., and Naidoo, G. C. 2000. Evaluation of perennial *Glycine* species for resistance to soybean fungal pathogens that cause Sclerotinia stem rot and sudden death syndrome. Crop Sci. 40: *545-549.*

Hartman, G.L., Wang, T.C., and Hymowitz, T. 1992. Source of resistance to soybean rust in perennial *Glycine* species. Plant Dis. 76:396-399.

Hymowitz, T. 1995. "I-D. Evaluation of wild perennial *Glycine* species and crosses for resistance to *Phakospora*," Proceedings of the Soybean Rust Workshop, Aug. 9-11, 1995, Sinclair et al. eds., National Soybean Research Laboratory Publication, No. 1, pp. 33-37.

Hymowitz, T., Hartman, G. L. 2015. History and growth of the soybean plant. Pages 1-4 in: Compendium of Soybean Diseases and Pests, G. L. Hartman, J. C. Rupe, E. F. Sikora, L. L. Domier, J. A. Davis and K. L. Steffey, eds. American Phytopathological Society, St. Paul.

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

Jiao, Y. Q., Vuong, T. D., Liu, Y., Meinhardt, C., Joshi, T., Cregan, P. B., Xu, D., Shannon, J. G., and Nguyen, H. T. 2015. Identification and evaluation of quantitative trait loci underlying resistance to multiple HG types of soybean cyst nematode in soybean PI 437655. Theor. Appl. Genet. 128:15-23.

Joos, D. K., Esgar, R. W, Henry, B. R., and Nafziger, E. D. 2013. Soybean Variety Test Results in Illinois-2013. Crop Sciences Special Report 2013-04. Univ. of Illinois, Urbana.

Kenworthy, W.J. 1989. Potential genetic contributions of wild relatives to soybean improvements. P.883-888. In: A.J. Pascale, ed. World Research Conference IV. Proc., Buenos Aires, 5-9 Mar. 1989. Argentina, Assoc. Soja, Buenos Aires.

Lim, S.M., and Hymowitz, T. 1987. Reaction of perennial wild species of genus *Glycine* to *Septoria glycines*. Plant Dis. 71:891-893.

Mignucci, J. S. and Chamberlain, D. W. 1978. Interactions of *Microsphaera diffusa* with soybeans and other legumes. Phytopathol. 68:169-173.

Mitchum, M. G., Wrather, J. A., Heinz, R. D., Shannon, J. G., and Danekas, G. 2007. Variability in distribution and virulence phenotypes of *Heterodera glycines* in Missouri during 2005. Plant Dis. 91:1473–1476.

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

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

Niblack, T. L., and Riggs, R. D. 2015. Soybean Cyst Nematode. Pages 100-104 in: Compendium of Soybean Diseases and Pests, G. L. Hartman, J. C. Rupe, E. F. Sikora, L. L. Domier, J. A. Davis and K. L. Steffey, eds. American Phytopathological Society, St. Paul.

Niblack T.L., Wrather J.A., Heinz R.D., and Donald P.A. 2003. Distribution and virulence phenotypes of *Heterodera glycines* in Missouri. Plant Dis. 87:929–932.

Nickell, C. D., Noel, G. R., Cary, T. R., Thomas, D. J., and Leitz, R.A. 1999. Registration of 'Ina' soybean. Crop Sci. 39:1533.

Niks, R. E., and Marcel, T. C. 2009. Nonhost and basal resistance: how to explain specificity? New Phytol. 182: 817-828.

Riggs, R.D. 1992. Host range. Pages 107-114 in Riggs, R.D. and Wrather, J.A. eds. Biology and Management of the Soybean Cyst Nematode. St. Paul, MN: American Phytopathological Society.

Riggs, R. D., and Schmitt, D. P. 1988. Complete characterization of race scheme of *Heterodera glycines*. J. Nematol. 20:392-395.

Riggs, R. D., Wang, S., Singh, R. J., and Hymowitz, T. 1998. Possible transfer of resistance to *Heterodera glycines* from *Glycine tomentella* to *Glycine max*. Supplement J. Nematol. 30:547-552.

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

Sherman-Broyles, S., Bombarely, A., Powell, A. F., Doyle J. L., Egan, A. N., Coate, J. E., Doyle, J. J.2014. The wild side of a major crop: soybean's perennial cousins from Down Under. Am. J. Bot. 101:1651-65.

Venkatesh, R., Harrison, S. K., and Riedel, R. M. 2000. Weed hosts of soybean cyst nematode (*Heterodera glycines*) in Ohio. Weed Technol. 14:156-160.

Wang, J., Niblack, T. L., Tremain, J. A., Wiebold, W. J., Tylka, G. L., Marett, C. C., Noel, G.R., Myers, O., and Schmidt, M. E. 2003. Soybean cyst nematode reduces soybean yield without causing obvious aboveground symptoms. Plant Dis. 87:623-628.

Young, L. D. 1990. Soybean germplasm evaluated for resistance to races 3, 5, and 14 of soybean cyst nematode. Crop Sci: 30:735-736.

Zhang, H., Li, C., Davis, E. L., Wang, J., Griffin, J. D., Kofsky, J. and Song, B. H. 2016. Genome-wide association study of resistance to soybean cyst nematode (*Heterodera glycines*) HG Type 2.5. 7 in wild soybean (*Glycine soja*). Front. Plant Sci. 7:1214.

Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, W., Yu, Y., Shu, L., Zhao, Y., Ma, Y. and Fang, C. 2015. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. Nature Biotechnol. 33:408-414.

CHAPTER 2. GENOME-WIDE ASSOCIATION STUDY AND GENOMIC PREDICTION ELUCIDATE THE GENETIC ARCHITECTURE OF SOYBEAN CYST NEMATODE RESISTANCE IN COMMON BEAN

ABSTRACT

Genome-wide association study (GWAS) was used to detect SNPs significantly associated with soybean cyst nematode (SCN, *Heterodera glycines* (HG)) resistance in the core collection of *Phaseolus vulgaris* and to make genomic predictions (GPs) of SCN resistance to two HG types. There were 84,416 SNPs identified in 363 common bean accessions. GWAS identified SNPs on chromosome (Chr) 1 that were significantly associated with resistance to HG type 2.5.7. These SNPs were in linkage disequilibrium with a gene cluster homologous to the three genes at *Rhg1* locus in soybean. A novel signal on Chr 7 was detected to be responsible for resistance to HG type 1.2.3.5.6.7. Furthermore, GP for resistance to two SCN HG types were highly predictive. Our study reported a high-quality SNPs set for the common bean collection, and we demonstrated that both GWAS and GP were effective strategies to study the genetic architecture of SCN resistance in common bean.

INTRODUCTION

Common bean (*Phaseolus vulgaris* L.) is one of the most important grain legumes in the human diet and a major protein source in many developing countries (Broughton et al. 2003). Common bean has two geographical and genetic pools, one of which is the Mesoamerica genetic pool domesticated in Mexico and another is the Andean genetic pool domesticated in Central and South America (McClean et al. 2004; Singh et al. 1991).

Common bean and soybean (*Glycine max* (L.) Merr.) are in the *Fabaceae* family, and are attacked by some of the same pathogens (Miles et al. 2007; Singh and Schwartz 2010; Souza et al. 2014). Soybean cyst nematode (SCN), *Heterodera glycines* (HG) Ichinohe, one of the most destructive soybean pathogens with production loss estimates averaged over the three-top soybean-producing countries exceeding 2% (Hartman et al. 2015), also infects *P. vulgaris*. For example, the Kidney bean variety "Clark" was shown to be a host for SCN HG type 0 as the variety supported juvenile growth, enlargement, molting, and female reproduction similar to the susceptible soybean cultivar "Amsoy 71" (Abawi and Jacobsen 1984). Other reports include *P. vulgaris* susceptibility and resistance to SCN. Out of 23 snap bean varieties inoculated with two SCN populations, only one resistant variety was found while all the other varieties supported equal or greater cyst production than the susceptible soybean cultivar "William 79" (Melton et al. 1985). A range of SCN female indices from 5 (on a black bean cultivar) to 117 (on a Kidney bean cultivar) was found when 24 dry bean varieties were inoculated with SCN HG type 0 (Poromarto and Nelson 2009).

The genetic architecture of SCN resistance in soybean has been intensively studied and identification of quantitative trait loci (QTL) have been used to develop SCN resistance in soybean (Diers and Arelli, 1999; Concibido et al., 2004). Two genetic resistance loci, *Rhg1* on chromosome (Chr) 18 and *Rhg4* on Chr 8 were repeatedly detected in bi-parental linkage mapping in different soybean studies (Bao et al. 2014; Cook et al. 2012; Liu et al. 2012). More recently, genome-wide association study (GWAS) has been widely used to identify SCN resistance-associated SNPs in soybean. The first study with 159 Chinese soybean accessions identified six simple sequence repeat (SSR) markers associated with SCN resistance (Li et al. 2011). The second study focused on SCN HG type 0 in 282 soybean breeding lines using the

universal soy linkage panel 1.0 SNP arrays, and discovered two previously identified QTLs, *Rhg1* and *FGAM1* (Vaghchhipawala et al. 2004) as well as a novel locus on the opposite end of Chr 18 to *Rhg1* (Bao et al. 2014). SNPs discovered in a third study overlapped six previously reported QTL (including the *Rhg1* and *Rhg4*) and eight novel QTLs were identified (Voung et al. 2015). Another GWAS study detected 19 signals significantly associated with SCN HG type 0 and HG type 1.2.3.5.7 among a collection of 440 diverse soybean landraces and elite cultivars. The SCN resistant loci, *Rhg1* and *Rhg4*, were identified together with three novel loci (Han et al. 2015). Recently, SNPs significantly associated with SCN race 1, 3 and 5 were also reported in the United States Department of Agriculture (USDA) Soybean Germplasm Collection (Chang et al. 2016a). One of these significant SNPs was in LD with five LRR-RLK genes on Chr 18 near to the *Rhg 1* locus, and another SNP (ss715638409) was in the coding region of an LRR-RLK gene on Chr 20 with three additional LRR-RLK genes in the LD region.

GWAS is suitable for identifying large-effect genetic loci that cause phenotypic variation. However, GWAS might miss a large proportion of total genetic variation caused by numerous loci of small-moderate effects, and this additional variation in SCN response can be captured by genomic prediction through fitting all genome-wide SNPs in a linear model assuming all the markers contribute to the phenotypic variation (Desta et al. 2014). However, when applying all markers to predict a phenotype, the number of predictor variables is greater than the number of observations. Collinearity is inevitable when too many variables are included in the model, resulting in over-fitting and instability of prediction model (Gianola 2013). To address this problem, several regularization approaches have been proposed for genomic prediction (GP) models. One of the models is ridge regression, which addresses the problems of ordinary least

squares by imposing a penalty on the size of coefficients. The ridge coefficients minimize a penalized residual sum of squares,

$$\min_{w} ||Xw - y||_2^2 + \alpha ||w||_2^2$$

where, $\alpha \geq 0$ is a complexity parameter that controls the amount of shrinkage: the larger the value of α, the greater the amount of shrinkage and thus the coefficients become more robust to collinearity. Ridge regression genomic prediction model has been implemented for many traits (Bao et al. 2014; Chang et al. 2016b; Lorenz et al. 2012; Rutkoski et al. 2014), and has dramatically increased the efficiency of crop improvement (Heffner et al. 2009).

GWAS has been applied on a number of traits in common bean. For example, a study for bacterial blight resistance used 132 SNPs in 469 common bean breeding lines (Shi et al. 2011) and identified 12 significant SNPs that were co-localized with previously reported QTL as well as two novel QTL. In another study, the genetic architecture of five agronomic traits were investigated using 233 amplified fragment length polymorphisms, 105 SNPs and 80 SSR markers in 66 common bean genotypes (Nemli et al. 2014). Nonetheless, there is no bi-parental mapping or GWAS focusing on SCN resistance in common bean. In order to identify SNPs that are associated with SCN resistance in common bean, the core collection of common bean was genotyped using genotyping-by-sequencing (GBS) and phenotyped against two SCN HG types that are commonly found in Illinois fields, the HG type 2.5.7 and the HG type 1.2.3.5.6.7. In addition to SCN resistance, two agronomic traits (seed coat color and seed weight) with known QTL were included in this study to validate the reliability of GWAS methodology. GP was applied to estimate the GEBVs of common bean accessions for resistance to two SCN HG types,

and prediction accuracies were evaluated using cross-validation. Our work may represent the first GWAS and GP for SCN resistance in common bean.

MATERIALS AND METHODS

Plant materials and DNA preparation

A total of 363 common bean accessions representing from the Mesoamerican and Andean gene pool were used in this study. 171 accessions of the Central/South American (CA/SA) core collection, and 191 accessions of the Mexico (MA) core collection (Brick et al. 2006; McClean et al. 2012), were requested from the USDA/ARS Western Regional Plant Introduction Station (Pullman, WA, USA), and the accession G19833 with reference genome (Schmutz et al. 2014) was obtained from the International Center for Tropical Agriculture and used in this study.

Genomic DNA was extracted from the emerging trifoliate leaf using a standard cetyltrimethylammonium bromide protocol (Doyle and Doyle, 1987). Genomic DNA was quantified in 96-well plates using PicoGreen (Invitrogen, Carlsbad, CA) and was normalized to 20 ng/µl. A total of 500 ng DNA of each accession in a 96-well plate was digested by HindIII and BfaI restriction enzymes (New England Biolabs, Ipswich, MA), and 0.1 µM A1 adaptor and 10 µM A2 adapters (IDT DNA) was used for ligation in each well. The genomic libraries were pooled and cleaned up using a PCR purification kit (Qiagen, Valencia, CA), followed by an amplification step for 12 cycles using Illumina primers and Phusion DNA polymerase (NEB). Average fragment size was estimated on a Bioanalyzer 2100 (Agilent, Santa Clara, CA) using a DNA1000 chip following by a second column-cleaning.

Genotyping-by-sequencing (GBS)

Pooled libraries were adjusted to 10 nmol and sequenced with 100-bp single-end reads in one lane of HiSeq2500 (Illumina, San Diego, CA). SNP calling was performed using Tassel5 GBS v2 variant calling pipeline IGST-GBS (Glaubitz et al. 2014; Sonah et al. 2013). All reads were trimmed to 64 bp at the 3' end to make sure each base has Phred score greater than 30, and the trimmed sequence were aligned to the non-masked reference genome of *P. vulgaris* G19833 *Pvulgaris* v1.0 obtained from Phytozome v11.0 (Glaubitz et al. 2014; Schmutz et al. 2014) using bowtie2 with the parameters -D 20 -R 3 -N 0 -L 20 -i S, 1, 0.50 (same as "very-sensitive" setting), which is computationally slower but more sensitive and more accurate than the default "sensitive" setting (Langmead and Salzberg 2012). Missing SNPs were imputed using BEAGLE version 4.1 (Browning and Browning 2013). Insertion-deletion polymorphisms (Indels), SNPs with minor allele frequency (MAF) less than 0.05, and SNPs with heterozygosity greater than 0.05 were excluded from GWAS and GP analysis.

Phenotyping for SCN resistance and two agronomic traits

The 363 common bean accessions along with a soybean variety Williams 82 were planted in polyvinyl chloride tubes (3 cm diameter x 15 cm deep) and 18 to 19 tubes were randomly inserted in a plastic container (20 cm diameter x 25 cm deep) filled with pasteurized torpedo sand. Tubes without germination were replaced with extra seedlings from containers. Each tube is an experiment unit, and each plant at one-week-old stage was inoculated with 1 ml suspension containing approximately 2000 eggs of a SCN HG type (HG 2.5.7 or HG 1.2.3.5.6.7). All plants were maintained in 28^{o}C water baths with 16 hours light in the greenhouse. Thirty-five days after inoculation, roots were washed and cysts were collected from each plant. Cysts were counted under a dissecting microscope, and the number of cysts on each plant was recorded. Cysts on

each common bean accession were counted and the female index (FI) was calculated by dividing the mean number of females that developed on a tested accession by the mean number of females on the susceptible check 'W82', multiplied by 100 (Niblack et al., 2002; Niblack et al., 2009). There were two replications in this experiment, and replication was achieved over time. Because of the complexity of seed coat color in common bean, only black, red and white seeds were included in GWAS, with black seeds rated as 2, red seeds rated as 1, and white seeds rated as 0. Seed weight data were obtained from the Germplasm Resources Information Network (GRIN) (www.ars-grin.gov), and the seed weight of each accession was represented by the weight of 100 randomly selected seeds of that accession. Box-Cox method was performed to transform non-normally distributed trait including SCN HG type 1.2.3.5.6.7 resistance and seed weight.

Genome-wide association study (GWAS) and Genomic Prediction (GP)

GWAS was performed using the R package "Genomic association and prediction integrated tool (GAPIT)" (Lipka et al. 2012). Principal component analysis (PCA) was done based on SNPs identified using GBS. A kinship matrix was calculated to determine relatedness among individuals (Zhang et al. 2010). A unified mixed linear model (MLM), which includes both kinship and principal component, was used in this study. The Bayesian information criterion (BIC)-was calculated in GAPIT to determine number of PCs that should be included in the model. All SNPs with false discovery rate (FDR) below 0.1 were reported. The R package "Ridge-regression best linear unbiased prediction (rrBLUP)" was applied to estimate SNP effects by solving the MLM through the restricted maximum likelihood (REML) method (Whittaker et al. 2000). The performance of the genomic prediction model was estimated using ten-fold cross validation. The 363 accessions were randomly partitioned into ten similar-sized subsets (nine subsets with 36 accessions each, and one subset with 39 accessions), with nine subsets as

training set and one subset as the validation set, and each subset has exactly one chance to be validation dataset. In each training process, the SNP effects were estimated for predicting the GEBVs of accessions in the validation dataset. The ten-fold cross-validation process was then repeated for 100 iterations, and the predicted GEBV was averaged over the 100 iterations. Prediction accuracy was calculated as the correlation between GEBVs and true phenotypic values. The effect of SNPs number on prediction accuracy was estimated by including different numbers of SNPs (1000, 5000, 20000, 50000) evenly distributed across the genome for GP.

RESULTS

Phenotypic analyses for SCN resistance, seed coat color, and seed weight

Greenhouse evaluations of the common bean core collection for resistance to SCN HG type 2.5.7 resulted in a range of FIs from 0.5 to 198.9 in a normal distribution (Fig. 2.1A). Only 16 accessions showed high resistance (FI below 10) to SCN HG type 2.5.7 and 54 accessions showed moderate resistance (FI between 10 to 30). On the other hand, 160 accessions had high resistance and 164 accessions had moderate resistance to SCN HG type 1.2.3.5.6.7. The FI to SCN HG type 1.2.3.5.6.7 was left skewed and Box-Cox transformation was applied to normalize the phenotype data (Fig. 2.1B). There were 19 accessions with white seed coat, 50 accessions with red seed coats, and 90 accessions with black seed coats. The seed weight of the 363 common bean accessions (weight of randomly selected 100 seeds) ranged from 2g to 91.6g, with approximately a normal distribution (Fig. 2.1C).

SNP calling and linkage disequilibrium (LD) decay analysis

Illumina sequencing yielded 264,276,230 raw reads. After quality control for raw reads, SNP calling using *P. vulgaris* G19833 as the reference genome, and SNP imputation, a total of 84,416

SNPs were obtained and these SNPs distributed over 11 chromosomes with an average of 7,674 SNPs per chromosome (Table 2.1). LD decay was estimated for each chromosome and ranged from 50 kb to 70 kb at a cutoff of correlation coefficients (r^2) equal to 0.2, and there were about 10 SNPs in one LD window.

Population structure

The population structure of the 363 common bean accessions was estimated by PCA using the 84,416 SNPs, and distinct subpopulations that matched with geographic origins were detected (Fig. 2.2A). The Mexico group and central American group had some overlap, but the southern American group clustered distinct from the other two groups, which is consistent with the two genetic pools theory of the common bean (Gepts et al. 1986; Koenig and Gepts 1989). On the other hand, accessions with different SCN resistance levels did not cluster into these distinct subgroups, indicating a mild confounding concern between subpopulations based on geographic origins and SCN resistance (Fig. 2.2B). BIC-based model selection also suggested that no principal component was required to control for population structure (Table 2.2). The kinship analysis with genetic relatedness among the 363 common bean accessions identified two clades, which is consistent with the prior knowledge of two genetic pools (Fig. 2.2B). Therefore, a unified MLM with a kinship matrix but no principal component was used for GWAS.

Genome-wide association study (GWAS) for seed coat color and seed weight

The GWAS for coat color and seed weight were compared to the results in the literature in order to validate our methodology. For seed coat color, the known locus *V* was mapped on linkage group 6 by several independent studies (Bassett 1997; McClean et al. 2002; McClean et al. 2002; Nodari et al. 1993). A random amplification of polymorphic DNA (RAPD) marker OD12800 on Chr 6 between 10480539bp to 10480584bp linked in coupling phase with the *V*

locus was reported (McClean et al. 2002), and it is in LD with the highly significant SNP detected in our study on Chr 6 9.6 Mb (Table 2.3; Fig. 2.3A and 2.3B).

For seed weight, GWAS identified 14 SNPs distributed over six regions on four different Chrs (Chr 2, Chr 3, Chr 7, and Chr 11) with FDR lower than 0.05 (Table 2.3; Fig. 2.3C and 2.3D). Previous linkage mapping studies for seed weight discovered QTL on linkage groups Chr 2, Chr 3 (Tsai et al. 1998), Chr 6, Chr 7, Chr 8, and Chr 11 using biparental populations (Blair et al. 2006; Park et al. 2000; Tsai et al. 1998). Most of the identified seed weight candidate genes belong to components of the cytokinin synthesis and multiple-component phosphorelay regulatory system (Schmutz et al. 2014). The three significant SNPs on Chr 2 (43,368,553bp, 43,379,956bp and 43,400,258bp) detected in our study are in LD with the candidate gene Phvul.002G282200 which was reported to be on Chr 2 between 44,603,605 bp and 44,608,648 bp; and the significant SNPs on Chr 3 at 5,204,703bp in our study is in LD with the candidate genes Phvul.003G041200 which is on Chr 3 between 4,582,905bp and 4,584,971 bp. The success in re-discovering previously reported QTLs for seed coat color and seed weight indicated the correctness and robustness of GWAS methodology.

Genome-wide association study (GWAS) for SCN resistance

For the first SCN HG type that we evaluated, the HG type 2.5.7, a genomic region on Chr 1 where four SNPs were identified with FDR below 0.05. Using a less stringent FDR cutoff at 0.1, SNPs located at two other regions (Chr 1 and Chr 9) were found (Table 2.3; Fig. 2.4A and 2.4C). The significant SNPs on Chr 1 explained about 5.9 % to 6.1 % of phenotypic variation, and additional 5.9 % and 5.3 % of phenotypic variation were explained by the two SNPs on other locations on Chr 1 and on Chr 9 (Table 2.3). Comparative genomic study mapped three genomic

clones (Bng122, Bng166, and Bng225) on common bean Chr 1 in the region near the *Rhg1* locus on soybean by testing random common bean clones as RFLP probes in soybean the putative bean genome location on Chr 1 was syntenic to the SCN resistance gene *Rhg1* in soybean by comparing RFLP marker Bng171a (Concibido et al. 1996). The *Rhg1* locus in soybean comprises these three genes: Glyma18g02580 (amino acid transporter), Glyma18g02590 (α-SNAP protein), and Glyma18g02610 (wound-inducible protein 12) (Cook et al. 2012). We blasted the three genes against the common bean genome, and the best hits for the three genes in the *P. vulgarius* genome are grouped together in a region from 50629261 bp to 50655828 bp on Chr 1. The hit for Glyma18g0258 was a hypothetical protein PHAVU_001G248000g on Chr 1 between 50653407 bp and 506558280 bp. The hit for Glyma18g02590 was a hypothetical protein PHAVU_001G247900g on Chr 1 between 50646068 bp to 50650097 bp. The hit for Glyma18g02610 was the hypothetical protein PHAVU_001G247700g on Chr 1 between 50629261 bp to 50630123 bp. Interestingly, the position of the three genes in the *P. vulgaris* genome are inverted from those in the soybean genome, and the significant SNPs on Chr 1 were in LD with the syntenic genomic region (Fig. 2.4B).

For the second HG type of SCN, the HG type 1.2.3.5.6.7, only one SNP on Chr 7 was detected below FDR of 0.1 (Table 2.3; Fig. 2.5A and 2.5B), and phenotypic variation explained by this SNP was estimated around 5.9 %. However, blast of the gene sequences at *Rhg 1* and *Rhg 4* locus did not return similar sequence match on Chr 7.

Genomic prediction (GP) for SCN resistance, seed coat color, and seed weight

Besides identifying SNPs associated with SCN resistance, seed coat color, and seed weight through GWAS, the effect of all the SNP markers were evaluated by GP models to predict SCN resistance, seed weight and seed color. The average prediction accuracy of the models estimated

by cross validation was 0.52, 0.41 and 0.82 for SCN HG2.5.7, HG 1.2.3.5.6.7 and seed weight. The prediction accuracy was not severely affected by marker size. For resistance to HG type 2.5.7, slight decrease in prediction accuracy was observed when number of markers were reduced 5000 and 1000 (Fig. 2.6A). For resistance to HG type .1.2.3.5.6.7, only the prediction accuracy with 1000 SNPs showed decreased prediction accuracy (Fig. 2.6B). The prediction accuracy for seed weight was not affected by number of SNPs (Fig. 2.6C), which indicated redundancy of the markers due to high level of LD.

DISCUSSION

In this study, 363 common bean accessions from the USDA Germplasm Collection were evaluated for their responses to two SCN HG types as well as two agronomic traits, seed coat color and seed weight. The identification of SNPs that are significantly associated with SCN resistance not only helps the genetic improvement of cultivars but also facilitates the identification of genes and the understanding of molecular mechanisms involved in the resistance process. In our study, GWAS identified novel SNPs associated with seed weight on Chr 7 and Chr 11, previously reported seed coat color gene locus *V* on Chr 6, and SNPs associated with SCN resistance to two SCN HG types. The significant SNPs identified for resistance to HG type 2.5.7 were in LD with the cluster of genes homologous to the *Rhg 1* locus in soybean (Concibido et al. 1996; Kelly et al. 2003). The genome blocks in *P. vulgaris* are conserved with the genomic region near the SCN resistance locus *Rhg1*, and the homologous genes in *P. vulgaris* are inverted with the three genes at *Rhg 1* locus in soybean.

Before soybean underwent a major duplication event 11 million years ago it diverged from a common bean ancestor (Schlueter et al. 2004). Comparative genomics studies reported 55

syntenic blocks between the two species (McClean et al. 2010) and it was shown that the linkage group D1 (Chr 1) of *P. vulgaris* was collinear with the top of linkage group G (Chr 18) of soybean (Concibido et al. 1996), which is consistent with our findings. Our findings support a gene cluster in common bean that is similar to the *Rhg 1* locus in soybean.

The advance in next generation sequencing technologies such as GBS in acquiring high quality and high density SNPs have not only enabled GWAS to achieve higher mapping resolution (Buckler and Thornsberry 2002; Yu and Buckler 2006) but also empowered GP to predict GEBVs for a quantitative trait in phenotype unknown accessions (Bao et al. 2014; Bernardo and Yu 2007; Chang et al. 2016a; Jannink et al. 2010; Lorenz et al. 2012; Rutkoski et al. 2014). Unlike marker-assisted selection, GP utilizes all the SNP information in the genome, which is expected to be more effective than marker assisted selection, especially for quantitative traits, in terms of capturing genetic variances with small effects (Heffner et al. 2009). GP provides the potential to increase genetic gain per generation through early selection for traits that are costly to phenotype. Phenotyping SCN resistance is a labor and time demanding process, thus implementation of genomic prediction in breeding for SCN resistance varieties might decrease the time and effort in phenotyping great number of crop accessions. Estimation of prediction accuracy for resistance to SCN HG 2.5.7 and HG 1.2.3.5.6.7 was 0.52 and 0.41, respectively, which is comparable to the prediction accuracy for SCN resistance in soybean (Bao et al. 2014). The prediction accuracy of GP for seed weight was as high as 0.82. Given the high prediction accuracy for both quantitative resistance and agronomic traits, GP holds great potential for common bean breeding programs.

In this study, we acquired high-density and high-quality SNPs for the 363 common bean accessions using GBS, and successfully identified SNPs associated with resistance to two SCN

HG types. Our result detected the SCN resistances for two HG types located on different locations of Chr 1, Chr 7, and Chr 9. The most significant peak on Chr 1associated with resistance to HG type 2.5.7 was syntenic to a previously reported SCN resistance locus in soybean (Concibido et al. 1996; Kelly et al. 2003). The result of our study provided the first insight into the genetic architecture of SCN resistance in common bean, and we are also the first to apply GP to predict SCN resistance, and seed weight for 363 common bean accessions.

TABLES

Table 2.1. Linkage disequilibrium (LD) decay estimated for different soybean chromosomes

Chr No.[1]	Chr size (kb)	No. of SNPs[2]	LD decay (kb)[3]
1	52183.5	8571	70
2	49033.7	8559	60
3	52218.6	8549	60
4	45793.2	8247	65
5	40237.5	7313	65
6	31973.2	8600	65
7	51698.4	6289	60
8	59634.6	9333	50
9	37399.6	5073	60
10	43213.2	7662	60
11	50203.6	9220	50

[1]Chr No. – Chromosome number
[2]Number of SNPs used in present study.
[3]LD decay at $r^2 = 0.2$**.**

Table 2.2. Bayesian information criterion (BIC)-based model selection for deciding the optimal number of PCs in the final model. Largest BIC value indicates best model.

A: BIC information for association mapping analysis for HG 2.5.7 SCN resistance

Number of PCs	BIC	log Likelihood Function Value
0	-2080.58	-2071.76
1	-2096.34	-2069.14
2	-2098.14	-2069.88
3	-2099.87	-2069.97
4	-2101.36	-2070.73
5	-2097.63	-2069.57

B: BIC information for association mapping analysis for HG 1.2.3.5.6.7 SCN resistance

Number of PCs	BIC	log Likelihood Function Value
0	-206.59	-207.75
1	-209.65	-209.18
2	-212.31	-210.36
3	-212.87	-210.76
4	-210.34	-209.79
5	-209.82	-209.23

Table 2.3. Significant SNPs detected in the genome-wide association study for four traits

Trait	Chr	SNP Position	P value	maf	R^2 of model without SNP	R^2 of model with SNP	FDR-adjusted *P* value
HG2.5.7 SCN resistance	1	18388403	1.02E-06	0.07	0.115	0.176	0.0304
	1	18388408	1.12E-06	0.07	0.115	0.176	0.0304
	1	18388378	1.36E-06	0.07	0.115	0.175	0.0304
	1	18388392	1.53E-06	0.068	0.115	0.174	0.0304
	9	35068146	1.80E-06	0.166	0.115	0.174	0.0304
	1	10061925	4.94E-06	0.071	0.115	0.168	0.0695
HG1.2.3.5.6.7 SCN resistance	7	44761605	9.57E-07	0.204	0.158	0.217	0.0808
Seed coat color	6	9601167	7.95E-08	0.144	0.417	0.537	0.0067
	2	17276368	6.89E-07	0.403	0.417	0.519	0.0171
	2	17278934	8.51E-07	0.413	0.417	0.517	0.0171
	9	19120733	1.45E-06	0.028	0.417	0.512	0.0171
	2	47163511	1.53E-06	0.106	0.417	0.512	0.0171
	2	17278931	1.64E-06	0.459	0.417	0.511	0.0171
	3	50861768	1.81E-06	0.163	0.417	0.51	0.0171
	3	50928462	1.81E-06	0.163	0.417	0.51	0.0171
	6	5131072	1.82E-06	0.234	0.417	0.51	0.0171
	2	17294139	3.25E-06	0.406	0.417	0.505	0.0275
Seed weight	3	5204703	6.24E-07	0.167	0.617	0.645	0.0232
	11	16852607	6.40E-07	0.065	0.617	0.645	0.0232
	2	43368553	3.09E-06	0.193	0.617	0.641	0.0232
	2	43379956	3.09E-06	0.193	0.617	0.641	0.0232
	2	43400258	3.09E-06	0.193	0.617	0.641	0.0232
	7	17891385	3.10E-06	0.186	0.617	0.641	0.0232
	3	3306736	3.30E-06	0.198	0.617	0.641	0.0232
	3	3306784	3.30E-06	0.198	0.617	0.641	0.0232
	3	3307933	3.30E-06	0.198	0.617	0.641	0.0232
	3	3307946	3.30E-06	0.198	0.617	0.641	0.0232
	3	3307948	3.30E-06	0.198	0.617	0.641	0.0232
	3	3335111	3.30E-06	0.198	0.617	0.641	0.0232
	2	32146945	3.91E-06	0.193	0.617	0.641	0.0235
	2	32146953	3.91E-06	0.193	0.617	0.641	0.0235

FIGURES

Fig. 2.1. Phenotypic analyses for SCN resistance, seed coat color, and seed weight. (A) Frequency distribution of the diverse panel of 363 common bean accessions for SCN response to HG 2.5.7. Vertical axis shows the number of individuals in a particular range of cyst count, and the horizontal axis represents the SCN cyst count. (B) Frequency distribution of the diverse panel of 363 common bean accessions for SCN response to HG 1.2.3.5.6.7. Cyst count of the HG 1.2.3.5.6.7 was transformed to the 0.26th power. Vertical axis shows the number of individuals in a particular range of transformed cyst count, and the horizontal axis represents the transformed SCN cyst count. (C) Frequency distribution of the 363 common bean accessions in different ranges of seed weight. Vertical axis shows the number of individuals in a particular range of seed weight, and the horizontal axis represents the seed weight.

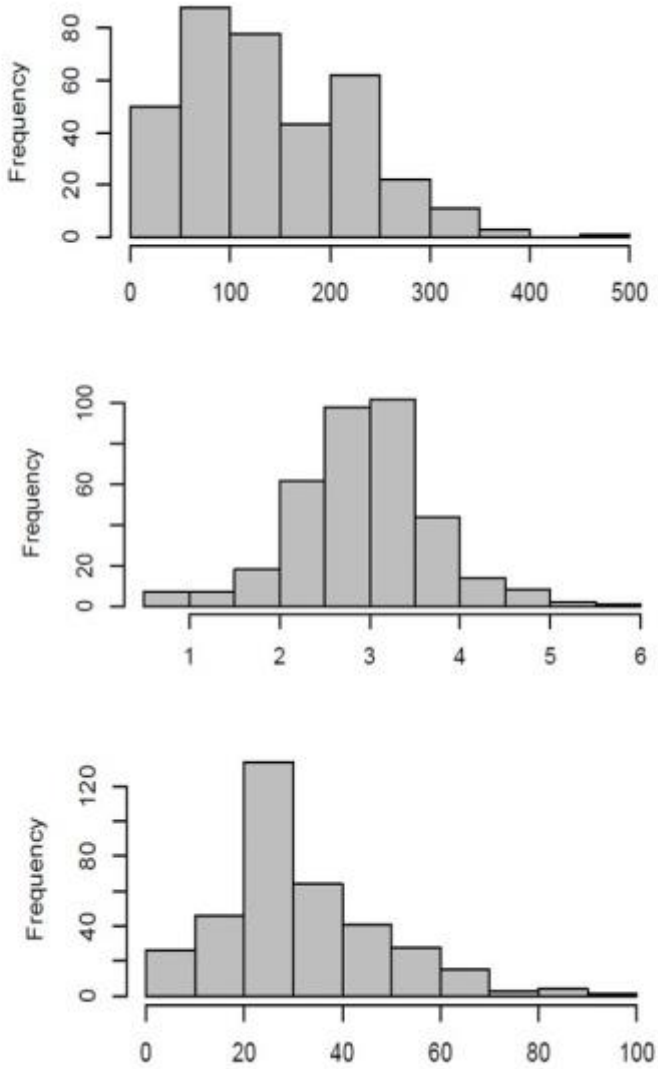


Fig. 2.2. Principal component analysis (PCA) and kinship matrix of the 363 common bean accessions genotyped with 84,416 SNPs. **(A) Genetic variation of the 363 common bean accessions explained by** the first three principal components calculated with the 84,416 SNPs. Different colors represent different origins (CA: Central America; MX: Mexico; SA: Southern America), and the principle components indicates distinct population structure. (B) and (C) **Genetic variation of the 363 common bean accessions explained by** the first three principal components calculated with the 84,416 SNPs. Different colors represent different levels of resistance (MR: moderately resistant; MS: moderately susceptible; R: resistant; S: susceptible) to HG type 2.5.7 and HG type 1.2.3.5.6.7, respectively. The results showed minor confounding effect between population structure and SCN resistance. (D) A kinship matrix of the 363 common bean accessions.

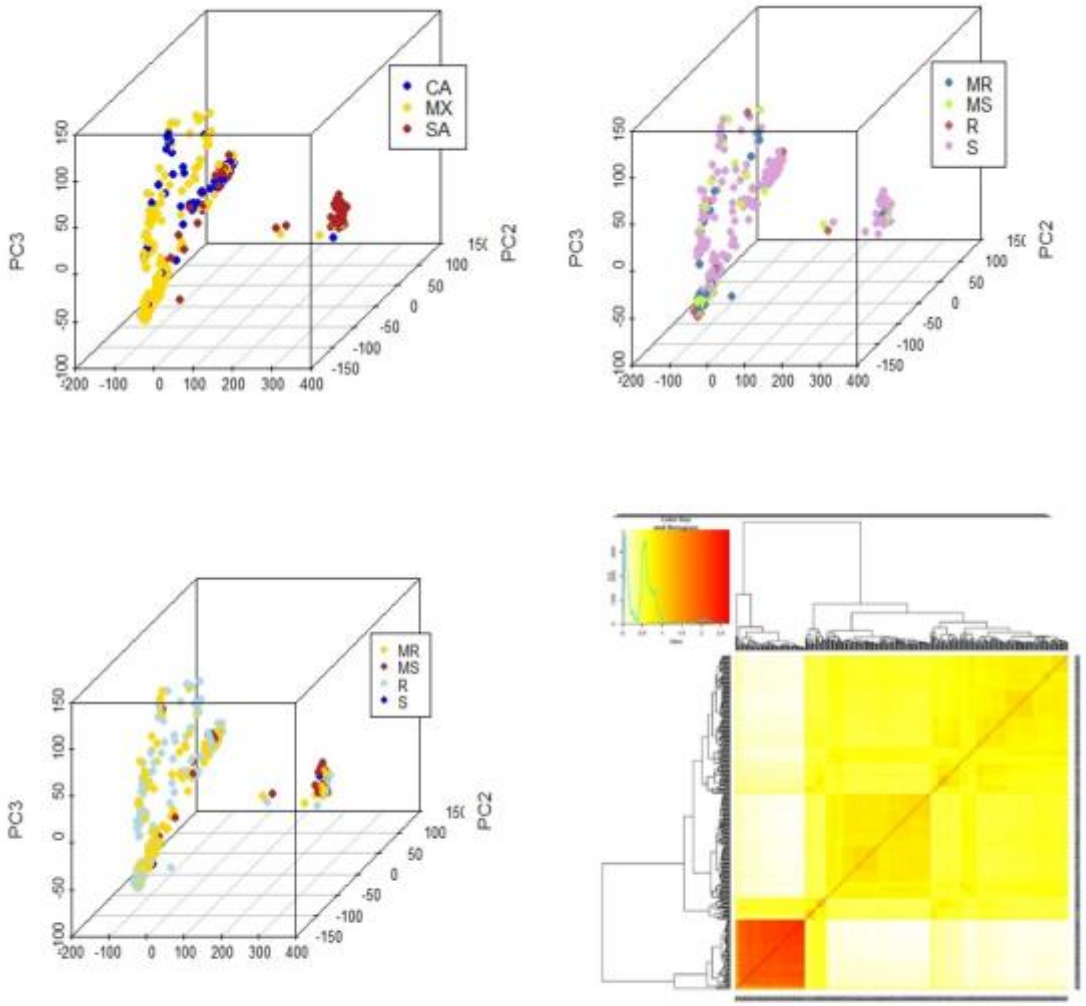


Fig. 2.3. GWAS for seed coat color and seed weight. (A) QQ plot for seed coat color (B) QQ plot for seed weight (C) Manhanttan plot for seed coat color (D) Manhanttan plot for seed weight.

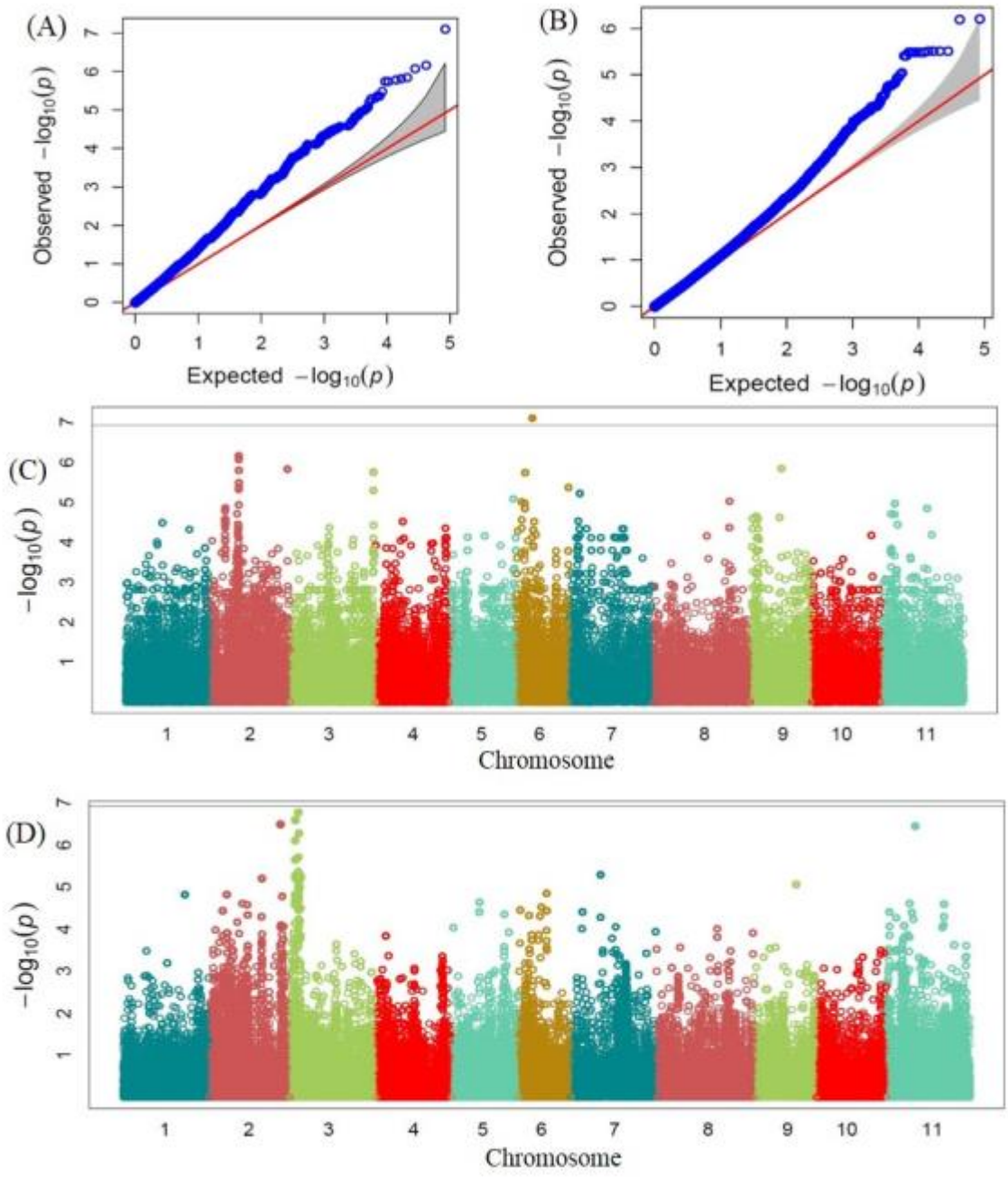


Fig. 2.4. GWAS for SCN resistance to HG type 2.5.7. (A) QQ plot for SCN resistance to HG type 2.5.7 (B) Pairwise LD displays of 15 SNPs located in the region surrounding SNPs detected by GWAS and the putative Rhg1 genes region on chromosome 1. The plot showed that the SNPs are in LD with the putative Rhg 1 gene. (C) Manhattan plot identified multiple significant SNPs on Chr. 1 and Chr.9.

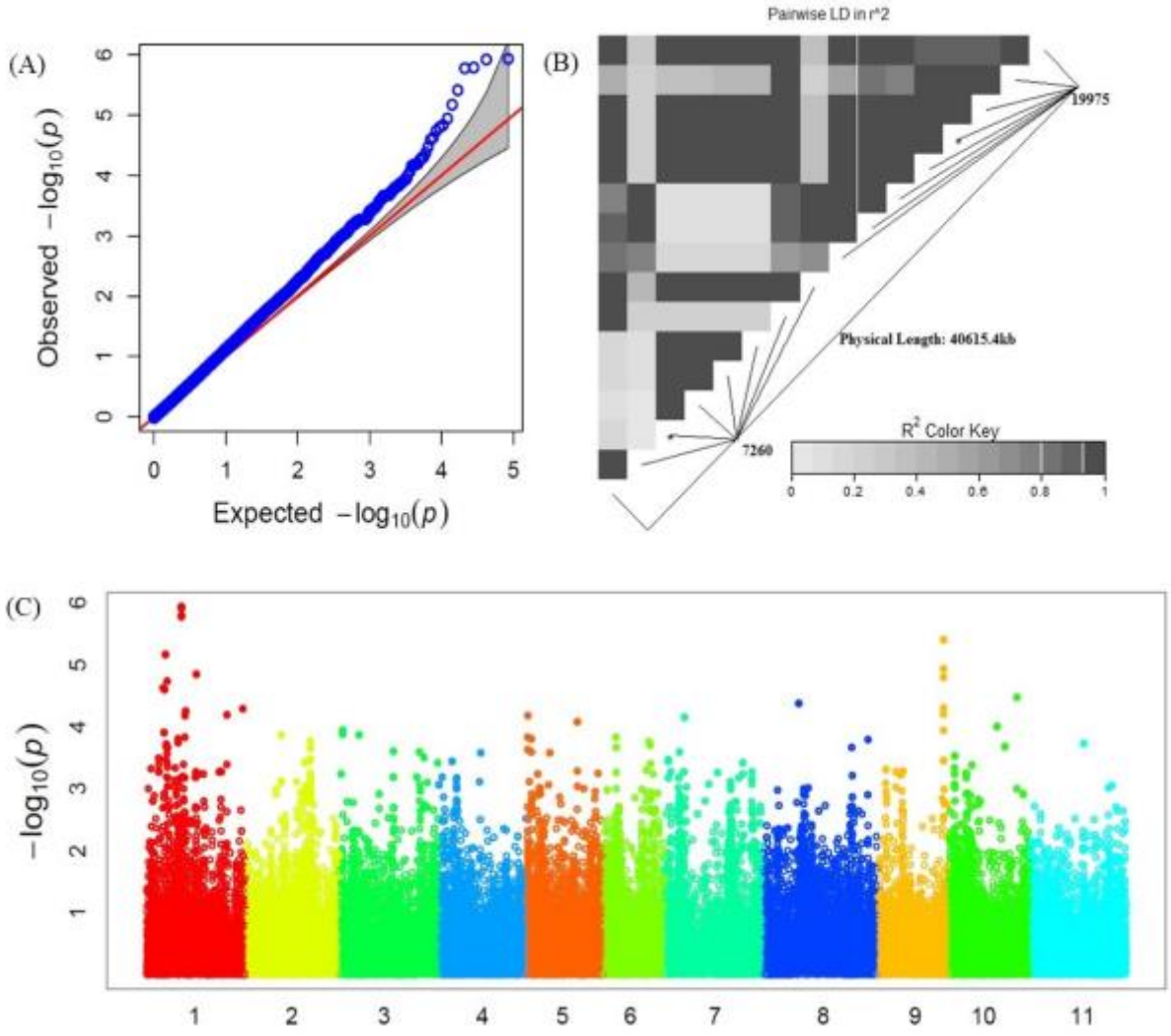


Fig. 2.5. GWAS for SCN resistance to HG type 1.2.3.5.6.7. (A) QQ plot for SCN resistance to HG type 1.2.3.5.6.7 (B) Manhattan plot identified one significant SNPs on Chr. 7.

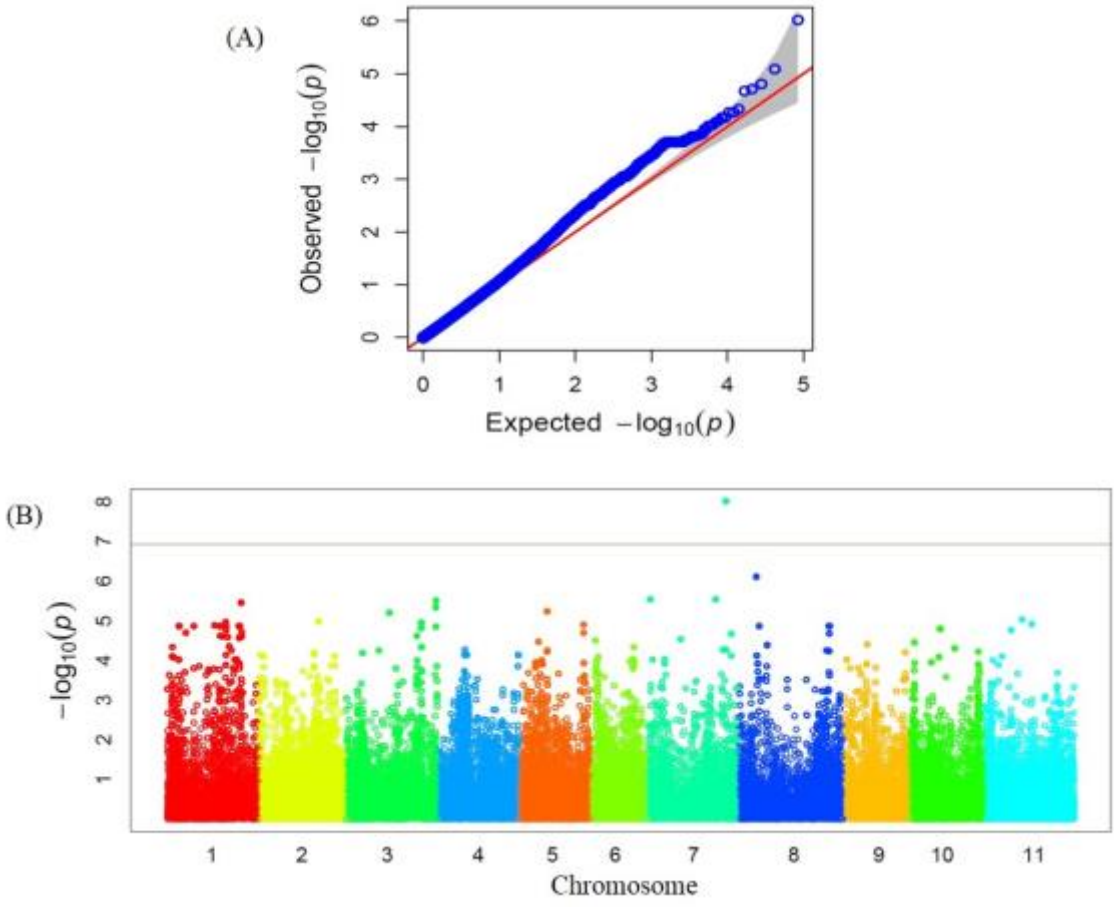
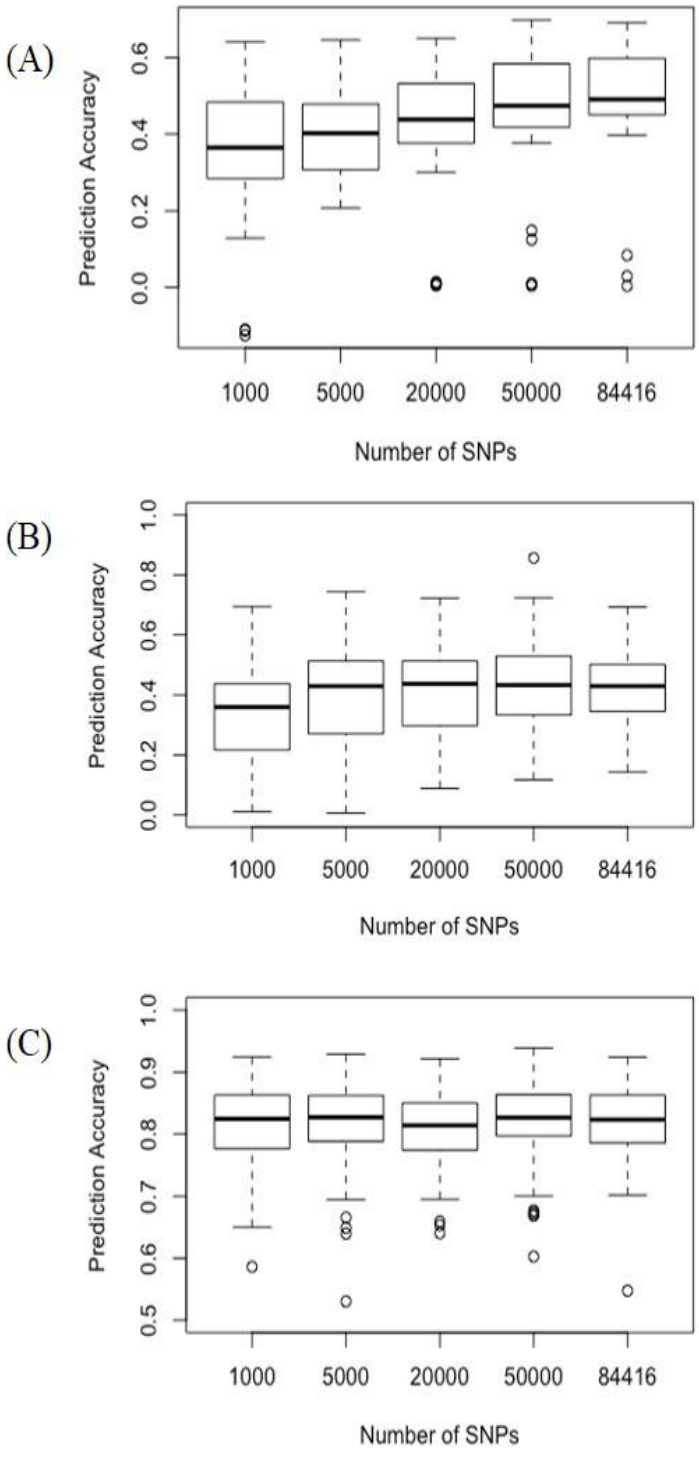


Fig. 2.6. Effect of SNP numbers used on prediction accuracy of genomic prediction. A) distribution of the prediction accuracy of 100 iterations of training process for predicting common bean resistance to SCN HG type 2.5.7. B) Distribution of the prediction accuracy of 100 iterations of training process for predicting common bean resistance to SCN HG type 1.2.3.5.6.7. C) A) distribution of the prediction accuracy of 100 iterations of training process for predicting common bean seed weight.



REFERENCES

Abawi, G.S., and B.J. Jacobsen.1984. Effect of initial inoculum densities of *Heterodera glycines* on growth of soybean and kidney bean and their efficiency as hosts under greenhouse conditions. Phytopathology 74:1470-1474.

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

Bassett, M.J. 1997. A new allele (Vwf) at the *V* locus for flower and seed coat color in common bean. J. Am. Soc. Hortic. Sci. 122:519-521.

Bernardo, R. and J. Yu. 2007. Prospects for genome-wide selection for quantitative traits in maize. Crop Sci. 47:1082-1090. doi:10.2135/cropsci2006.11.0690

Brick, M.A., P.F. Byrne, H.F. Schwartz, J.B. Ogg, K. Otto, A.L. Fall, and J. Gilbert. 2006. Reaction to three races of Fusarium wilt in the core collection. Crop Sci. 46:1245-1252.

Broughton, W.J., G. Hernandez, M. Blair, S. Beebe, P. Gepts, and J. Vanderleyden. 2003. Beans (*Phaseolus spp*.)–model food legumes. Plant Soil 252:55-128. doi:10.2135/cropsci2005.06-0102

Browning, B. L. and S.R. Browning. 2013. Improving the accuracy and efficiency of identity by descent detection in population data. Genetics 194:459-71. doi:10.1534/genetics.113.150029

Buckler, E.S. and J.M. Thornsberry. 2002. Plant molecular diversity and applications to genomics. Curr. Opin. Plant Biol. 5:107-111.

Chang, H.X., P.J. Brown, A.E. Lipka, L.L. Domier, and G.L. Hartman. 2016a. Genome-wide association and genomic prediction identifies associated loci and predicts the sensitivity of Tobacco ringspot virus in soybean plant introductions. BMC genomics 17:1.doi: 10.1186/s12864-016-2487-7

Chang, H.X., A. Lipka, L. Domier, and G.L. Hartman. 2016b. Characterization of disease resistance loci in the usda soybean germplasm collection using genome-wide associations. Phytopathology (ja).

Cook, D.E., T.G. Lee, X. Guo, S. Melito, K. Wang, A.M. Bayless, J. Wang, T.J. Hughes, D.K. Willis, T.E. Clemente, and B.W. Diers. 2012. Copy number variation of multiple genes at Rhg1 mediates nematode resistance in soybean. Science 338:1206-1209. doi:10.1126/science.1228746

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

Concibido, V.C., N. Young, D. Lange, R. Denny, D. Danesh, J. Orf. 1996. Targeted comparative genome analysis and qualitative mapping of a major partial-resistance gene to the soybean cyst nematode. Theoret. Appl. Genet. 93:234–241. doi:10.1007/BF00225751

Desta, Z.A. and R. Ortiz. 2014. Genomic selection: genome-wide prediction in plant improvement. Trends Plant Sci. 19: 592-601.

Diers, B.W., P.R. Arelli. 1999. Management of parasite nematodes of soybean through genetic resistance. In Proceedings World Soybean Research Conferene VI, edited by Kauffmann, H.E., pp.300-306. Superior Printing, Champaign, Illinois, USA.

Doyle, J. and J.L. Doyle. 1987. Genomic plant DNA preparation from fresh tissue-CTAB method. Phytochem. Bull. 19:11-15.

Gepts, P. and F.A. Bliss. 1986. Phaseolin variability among wild and cultivated common beans (*Phaseolus vulgaris*) from Colombia. Econ. Bot. 40: 469-478.
Gianola, D. 2013. Priors in whole-genome regression: The Bayesian alphabet returns. Genetics 194:573–596. doi:10.1534/genetics.113.151753
Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire, Q. Sun, and E.S. Buckler. 2014. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PLoS One 9:e90346.
Han, Y., X. Zhao, G. Cao, Y. Wang, Y. Li, D. Liu, W. Teng, Z. Zhang, D. Li, L. Qiu, and H. Zheng. 2015. Genetic characteristics of soybean resistance to HG type 0 and HG type 1.2. 3.5.7 of the cyst nematode analyzed by genome-wide association mapping. BMC genomics 16:1. doi: 10.1186/s12864-015-1800-1
Hartman, G. L. 2015. Worldwide importance of soybean pathogens and pests. Pages 4-5 in: Compendium of Soybean Diseases and Pests, G. L. Hartman, J. C. Rupe, E. F. Sikora, L. L. Domier, J. A. Davis and K. L. Steffey, eds. American Phytopathological Society, St. Paul.
Heffner, E.L., M.E. Sorrells, and J.L. Jannink. 2009. Genomic selection for crop improvement. Crop Sci. 49:1-12.
Jannink, J.L., Lorenz, A.J. and Iwata, H., 2010. Genomic selection in plant breeding: from theory to practice. Brief. Funct. Genomics 9:166-177. doi: https://doi.org/10.1093/bfgp/elq001
Kelly, J.D., P. Gepts, P.N. Miklas, D.P. Coyne. 2003. Tagging and mapping of genes and QTL and molecular marker-assisted selection for traits of economic importance in bean and cowpea. Field Crops Res. 82:135–154.
Koenig, R. and P. Gepts. 1989. Allozyme diversity in wild Phaseolus vulgaris: further evidence for two major centers of genetic diversity. Theor. Appl. Genet. 78:809-817. doi:10.1007/BF00266663
Langmead, B. and S.L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods. 9(4):357-359. doi:10.1038/nmeth.1923
Lorenz, A.J., K.P. Smith, and J-L. Jannink. 2012. Potential and optimization of genomic selection for Fusarium head blight resistance in six-row barley. Crop Sci. 52:1609-1621. doi:10.2135/cropsci2011.09.0503
Liu, S., P.K. Kandoth, S.D. Warren, G. Yeckel, R. Heinz, J. Alden, C. Yang, A. Jamai, T. El-Mellouki, P.S. Juvale, and J. Hill. 2012. A soybean cyst nematode resistance gene points to a new mechanism of plant resistance to pathogens. Nature 492:256-260. doi:10.1038/nature11651
Li, Y.H., M.J. Smulders, R.Z. Chang, and L.J. Qiu. 2011. Genetic diversity and association mapping in a collection of selected Chinese soybean accessions based on SSR marker analysis. Conserv. Genet. 12:1145-1157. doi:10.1007/s10592-011-0216-y
Lipka, A. E., F. Tian, Q.S. Wang, J. Peiffer, M. Li, P.J. Bradbury, M.A. Gore, E.S. Buckler, and Z.W. Zhang. 2012. GAPIT: genome association and prediction integrated tool. Bioinformatics 28:2397-2399. doi: 10.1093/bioinformatics/bts444
McClean, P.E., J. Kami, and P. Gepts. 2004. Genomics and genetic diversity in common bean. Legume Crop Genomics 60:82.
McClean, P.E., R.K. Lee, C. Otto, P. Gepts, and M.J. Bassett. 2002. Molecular and phenotypic mapping of genes controlling seed coat pattern and color in common bean (*Phaseolus vulgaris* L.). J. Hered. 93:148-152. doi: 10.1093/jhered/93.2.148

McClean, P.E., S. Mamidi, M. McConnell, S. Chikara, and R. Lee. 2010. Synteny mapping between common bean and soybean reveals extensive blocks of shared loci. BMC genomics 11:1. doi: 10.1186/1471-2164-11-184

McClean, P.E., J. Terpstra, M. McConnell, C. White, R. Lee, and S. Mamidi. 2012. Population structure and genetic differentiation among the USDA common bean (*Phaseolus vulgaris* L.) core collection. Genet. Resour. Crop Evol.59:499-515. doi:10.1007/s10722-011-9699-0

Melton, T.A., G.R. Noel, B.J. Jacobsen, and D.J. Hagedorn. 1985. Comparative host suitabilities of snap beans to the soybean cyst nematode (*Heterodera glycines*). Plant Dis. 69: 119-122.

Miles, M.R., M. A. Pastor-Corrales, G. L. Hartman, and R. D. Frederick, R.D. 2007. Differential response of common bean cultivars to *Phakopsora pachyrhizi*. Plant Dis. 91**:** 698-704.

Nemli, S., T.K. Asciogul, H.B. Kaya, A. Kahraman, D. Eşiyok, and B. Tanyolac. 2014. Association mapping for five agronomic traits in the common bean (*Phaseolus vulgaris* L.). J. Sci. Food Agr. 94:3141-3151. doi: 10.1002/jsfa.6664

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

Niblack, T., G.L. Tylka, P. Arelli, J. Bond, B. Diers, P. Donald, J. Faghihi, V. Ferris, K. Gallo, and R.D. Heinz. 2009. A standard greenhouse method for assessing soybean cyst nematode resistance in soybean: SCE08 (standardized cyst evaluation 2008). Plant Health Prog. 10.1094/PHP-2009-0513-01-RV.

Nodari, R.O., S.M. Tsail, R.L. Gilbertson, and P. Gepts. 1993. Towards an integrated linkage map of common bean 2. Development of an RFLP-based linkage map. Theor. Appl. Genet. 85:513-520. doi:10.1007/BF00220907

Poromarto, S.H., and B.D. Nelson. 2009. Reproduction of soybean cyst nematode on dry bean cultivars adapted to North Dakota and northern Minnesota. Plant Dis. 93:507-511.

Rutkoski, J.E., J.A. Poland, R.P. Singh, J. Huerta-Espino, S. Bhavani, H. Barbier, M.N. Rouse, J.L. Jannink, and M.E. Sorrells. 2014. Genomic selection for quantitative adult plant stem rust resistance in wheat. Plant Gen. 7(3). doi:10.3835/plantgenome2014.02.0006

Schlueter, J.A., P. Dixon, C. granger, D. Grant, L. Clark, J.J. Doyle, R.C. Shoemaker. 2004. Mining the EST databases to determine evolutionary events in the legumes and grasses. Genome 47:868–876. doi: 10.1139/g04-047.

Schmutz, J., P.E. McClean, S. Mamidi, G.A. Wu, S.B. Cannon, J. Grimwood, J. Jenkins, S. Shu, Q. Song, C. Chavarro, and M. Torres-Torres. 2014. A reference genome for common bean and genome-wide analysis of dual domestications. Nature Genet. 46:707-713. doi:10.1038/ng.3008

Shi, C., A. Navabi, and K. Yu. 2011. Association mapping of common bacterial blight resistance QTL in Ontario bean breeding populations. BMC Plant Biol. 11:1. doi: 10.1186/1471-2229-11-52

Singh, S.P., P. Gepts, and D.G. Debouck. 1991. Races of Common Bean (*Phaseolus vulgaris*, Fabaceae). Econ. Bot. 45:379-396. doi:10.1007/BF02887079

Singh, S. P., and H.F. Schwartz. 2010. Breeding common bean for resistance to diseases: A review. Crop Sci. 50:2199–2223. doi:10.2135/cropsci2009.03.0163

Sonah, H., M. Bastien, E. Iquira, A. Tardivel, G. Légaré, et al. 2013. An improved genotyping by sequencing (GBS) approach o ering increased versatility and e ciency of SNP discovery and genotyping. PLoS ONE 8(1):e54603. doi:10.1371/journal.pone.0054603

Souza, T.L.P.O., S.N. Dessaune, M.A. Moreira, and E.G. Barros. 2014. Soybean rust resistance sources and inheritance in the common bean (*Phaseolus vulgaris* L.). Genet. Mol. Res. 13: 5626.

Tsai, S.M., R.O. Nodari, D.H. Moon, L.E.A. Camargo, R. Vencovsky, P. Geps. 1998. QTL mapping for nodule number and common bacterial blight in *Phaseolus vulgaris* L. Plant Soil 204:135–145.

Vaghchhipawala, Z.E., J.A. Schlueter, R.C. Shoemaker, and S.A. Mackenzie. 2004. Soybean FGAM synthase promoters direct ectopic nematode feeding site activity. Genome 47:404–413.

Vuong, T.D., H. Sonah, C.G. Meinhardt, R. Deshmukh, S. Kadam, R.L. Nelson, J.G. Shannon, and H.T. Nguyen. 2015. Genetic architecture of cyst nematode resistance revealed by genome-wide association study in soybean. BMC Genomics 16:1. doi: 10.1186/s12864-015-1811-y

Whittaker, J.C., R. Thompson, and M.C. Denham. 2000. Marker-assisted selection using ridge regression. Genet. Res. 75:249-252.

Yu, J. and E.S. Buckler. 2006. Genetic association mapping and genome organization of maize. Curr. Opin. Biotechnol.17:155-160.

Zhang, Z., E. Ersoz, C.Q. Lai, R.J. Todhunter, H.K. Tiwari, M.A. Gore, P.J. Bradbury, J. Yu, D.K. Arnett, J.M. Ordovas, and E.S. Buckler. 2010. Mixed linear model approach adapted for genome-wide association studies. Nature Genet. 42:355–360. doi:10.1038/ng.546

CHAPTER 3. PREDICTION OF SHORT-DISTANCE AERIAL MOVEMENT OF *PHAKOPSORA PACHYRHIZI* UREDINIOSPORES USING MACHINE LEARNING[1]

ABSTRACT

Dispersal of urediniospores by wind is the primary means of spread for *Phakopsora pachyrhizi*, the cause of soybean rust. Our research focused on the short distance movement of urediniospores from within the soybean canopy and up to 61 m from field-grown soybean rust infected plants. Environmental variables were used to develop and compare models including the least absolute shrinkage and selection operator regression, zero-inflated Poisson/regular Poisson regression, random forest, and neural network to describe deposition of urediniospores collected in passive and active traps. All four models identified distance of trap from source, humidity, temperature, wind direction, and wind speed as the five most important variables influencing short distance movement of urediniospores. The random forest model provided the best predictions, explaining 76.1% and 86.8% of the total variation in the passive- and active-trap datasets, respectively. The prediction accuracy based on the correlation coefficient (r) between predicted values and the true values were 0.83 ($P < 0.0001$) and 0.94 ($P < 0.0001$) for the passive- and active-trap datasets, respectively. Overall, we found that using multiple machine learning techniques identified the most important variables to make the most accurate predictions of movement of *P. pachyrhizi* urediniospores found a short distance from the source.

[1]This chapter was published on Phytopathology (ISSN:0031-949X) and reprinted from Wen, L., Bowen, C.R, and Hartman, G.L. 2017. Prediction of short-distance aerial movement of *Phakopsora pachyrhizi* urediniospores using machine learning. Phytopathology. (ja). doi: 10.1094/PHYTO-04-17-0138-FI.

INTRODUCTION

Phakopsora pachyrhizi Syd. is the fungal pathogen that causes rust on soybean [*Glycine max* (L.) Merr.] and many other hosts (Hartman et al. 2015). Like many other fungi that cause plant rusts, *P. pachyrhizi* produces copious amounts of urediniospores that are readily wind-blown. Since the transoceanic spread of soybean rust from Asia to Hawaii, Africa, South America, and North America during the past two decades (Hartman et al. 2015), various models have been developed to predict *P. pachyrhizi* urediniospores dispersal from one location to another. In November 2004, soybean rust was reported for the first time in the continental United States in the state of Louisiana (Schneider et al. 2005). Based on long-distance transport models, *P. pachyrhizi* urediniospores may have been moved to the continental United States on winds associated with Hurricane Ivan (Isard et al. 2005), which occurred about two months before the discovery of soybean rust from Louisiana.

The discovery of soybean rust in the continental United States brought about much research activity due to the potential of the pathogen to devastate soybean production (Kelly et al. 2015). Yield losses of up to 80% were reported in experimental plots in Taiwan (Hartman et al. 1991), 30 to 75% in Brazil (Yorinori et al. 2005), and up to 31, 40, and 27% in Paraguay, South Africa, and Zimbabwe, respectively, (Miles et al. 2007). A network of sentinel plots was established to monitor the occurrences of the disease in the United States and information on its distribution was made available via the Pest Information Platform for Extension and Education (ipmPIPE) (Isard et al. 2006). In the continental United States, *P. pachyrhizi only* survives winters on kudzu and overwintering volunteer soybeans below the frost line (Kelly et al. 2015; Sikora et al. 2014; Yang 2006), and perhaps throughout the Caribbean, causing a bottleneck and often slowing development of rust in the South each spring (Mundt et al. 2013). Data collected through the

sentinel plot network shows that rust occurred throughout the southern United States every year since 2004, but the spread of rust to the northern soybean-growing regions has been less common and, when it occurs, it develops late in the season (USDA 2017). Most commonly soybean rust has spread locally (within and among adjacent soybean fields), at a rate of 8.8 and 10.4 km per day in soybean in 2005 and 2006, respectively (Christiano and Scherm 2007).

The dispersal of *P. pachyrhizi* urediniospores is an essential component of the spread of the pathogen, and occurs at all scales ranging from within the plant canopy to an entire hemisphere. The transport of urediniospores on different regional scales is important for disease modeling and forecasting. There is potential for *P. pachyrhizi* urediniospores to be lifted upward into the atmosphere for long-distance transport through rapidly moving air masses, similar to what may have occurred with Hurricane Ivan (Isard et al. 2005). Anticipating the spread of *P. pachyrhizi* urediniospores in the United States, simulation-based mathematical models were developed to assess the long-distance movement of *P. pachyrhizi* urediniospores from the southern to the northern United States (Isard et al. 2005; Isard et al. 2007; Isard et al. 2011; Krupa et al. 2006; Pan et al. 2006). One model used a climate-dispersion integrated model system to simulate long-distance and long-term movement of *P. pachyrhizi* urediniospores one month in advance of an epidemic (Pan et al. 2006). This aerobiological model is based on the Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model which simulates single trajectories of air parcels and deposition of particles originating from a source geographical location and time of year (NOAA 2006). Another model is the Soybean Rust Aerobiology Prediction System (SRAPS) (Isard et al. 2005), which simulates basic spore dispersal processes stated by Aylor (1999). These processes include spore production, release or removal from a substrate, escape from the canopy, transport and dilution by wind, loss of viability during transport, removal from

the atmosphere by precipitation, deposition on host tissue, and impact in a destination area (Isard and Gage 2001). Previous models that predict the long-distance movement of soybean rust were configured in a modular format to include all of the stages in the aerobiology process. For example, the SRAPS (Isard et al. 2005) includes the modules of spore release and canopy escape in source areas, mortality due to exposure to solar radiation during atmospheric transport, and wet deposition in destination regions using the National Center for Environmental Prediction–Department of Energy Reanalysis 2 data set (Kalnay et al. 1996, Kanamitsu et al. 2002).

The environmental variables governing each step of the aerobiological pathway of *P. pachyrhizi* urediniospores are important components in spore transport modeling, although there are gaps in knowledge about what influences the release and escape of rust urediniospores from the canopy to the atmosphere (Issard et al., 2005). The release and movement of *P. pachyrhizi* urediniospores from the canopy is influenced by wind and turbulence, the canopy structure, and location of the released spores within the canopy. Precipitation may also play a role in facilitating dislodgement of rust urediniospores from the uredinia (Del Ponte et al. 2006b). After escaping the soybean canopy, turbulent diffusion and wind shear dilute the spores that are transported by airflows, and weather variables such as solar radiation, temperature, and relative humidity affect the survival of spores and dry deposition by wind and turbulence (Gregory 1973) and wet deposition by precipitation (Aylor 1986). After spore deposition on soybean leaves, various variables such as leaf wetness, temperature, and relative humidity, will further affect the establishment of rust (Desborough 1984; Melching et al. 1989).

The escape of *P. pachyrhizi* urediniospores from the substrate to the air above the canopy is a key component in the short- and long-distance movement modeling. However, since the short distance movement of soybean rust is poorly understood, parameters involved in this process,

including strength of inoculum source, escape rate of urediniospores (from infected leaves and the canopy), survival, and deposition rates, were often estimated based on previous knowledge and experience (Isard et al. 2005; Pan 2006; Yang 2006). Some quantitative data has been reported for the amount of urediniospores generated from soybean, susceptible kudzu, and resistant kudzu canopy in Florida; this was estimated at 6.7×10^{12}, 4.4×10^{12}, and 7×10^{11} urediniospores per day, respectively (Kelly et al. 2015). The application of SRAPS makes multiple assumptions, including that 25% of the heavily-infected source plant releases 6 million urediniospores per day, 33% of which are released during the optimal transport period and that 15% of them escape from the soybean canopy (Isard et al. 2005). Considering the varying local production and availability of *P. pachyrhizi* urediniospores, the accurate estimate of the short distance movement of rust urediniospores may be crucial to risk assessments and seasonal soybean rust forecasting.

Understanding how environmental factors influence the early stages of aerobiological process provides useful information for modeling of short distance movement from field to field and long-distance spore movement from within a continent or from continent to continent. One model simulated urediniospores escape from the plant canopy through wind speed to predict urediniospores escape from the boundary layer of the canopy (Andrade et al. 2005). Along with wind speed, the release of *P. pachyrhizi* urediniospores also depends on other environmental factors like humidity, leaf wetness, precipitation, solar radiation, and temperatures (Del Ponte et al. 2006b; Isard et al. 2005; Isard et al. 2006a). However, direct evidence of the relationships among these factors and the short distance transport of rust urediniospores is not known.

Predicting the short distance movement of rust spores with environmental variables include data that is complex with multicollinearity. Unlike traditional regression methods, machine

learning methods are appropriate for complex data including non-linear relationships between predictors and a response and are also able to process noisy data with multicollinearity (Recknagel 2001; Garzón et al. 2006).

The absolute shrinkage and selection operator (LASSO) is a well-established method that reduces the variance of the coefficient estimates by shrinking the coefficients, some to zero, and has been used for both feature selection and multicollinearity (Oyeyemi et al. 2015). Poisson regression is a commonly used model for analyzing count variables. The zero-inflated Poisson regression model is a modification of the regular Poisson regression model, but it allows for excessive counts of zero in the data. The central idea of the zero-inflated Poisson regression model is that the data comes from two regimes; one is from counts of zero and the other follows a standard Poisson distribution (Lambert 1992). Random forest model is a tree-based method, in which the data space is recursively partitioned based on the value of one of the predictor variables, such that the observations within a partition become more and more homogeneous (Matsuki et al. 2016). Building multiple decision trees using random samples of data points for each tree and random samples of predictors at each split point, random forest can provide more accuracy than those of any single tree and prevent overfitting (Breiman 2001). Neural network models are highly sophisticated pattern recognition systems capable of learning complex relationships. During the neural network learning process, input variables are weighted in various combinations, summed, and passed on to multiple layers. This combination of simple calculations results in the ability of neural networks to learn complicated non-learner relationships (Livingstone et al. 1997).

The objective of our study was to develop empirical models to predict the short distance transport of *P. pachyrhizi* urediniospores based on environmental variables, using urediniospores

count data collected from passive and active spore traps at four environments in the southern United States. Our study compared models using the LASSO, zero-inflated Poisson/regular Poisson model, random forest, and neural network to predict the short distance transport of *P. pachyrhizi* urediniospores with environmental variables.

MATERIALS AND METHODS

Environments. Rust epidemics were monitored in soybean fields at Fairhope, Alabama from 11-19 August 2005, Attapulgus, Georgia from 4-11 October 2005, and Quincy, Florida from 5-12 October and 2-12 November 2005. Soybeans at Fairhope and Attapulgus were at growth stages (GS) R6-R7 and R7-R8, respectively, and at Quincy-1 at GS R2 and Quincy-2 at GS R4-R6.

At each environment, a Watchdog 700 weather station (Spectrum Technologies, Inc., Plainfield, Illinois 60585) logged average temperature (˚C), relative humidity (%), wind speed (km/h), wind direction (˚), solar insolation (watts/m^2), precipitation (mm), and leaf wetness (scale of 0 to 15, where 0 = dry and 15 = wet) at 60 min intervals.

Passive trap urediniospore collection. A network of 16 passive wind-directional spore traps, modeled after the aeroconiscope (Maddox 1870; Maddox 1871), formed two concentric rings around the soybean rust source fields (Fig. A.1-A.2). Traps were positioned at 15 m and 61 m from the edges of diseased plants in plots at 0° (north), 45° (northeast), 90° (east), 135° (southeast), 180° (south), 225° (southwest), 270° (west), and 315° (northwest) orientations. Each trap consisted of a 30-cm length of 10.2 cm diameter PVC pipe with a 25 cm x 25 cm x 2 mm tail of plastic sheeting fitted into a slit at one end and oriented perpendicular to the ground. Traps behaved like wind vanes as moving air on the surface of the tail caused traps to rotate on swivel

mechanisms orienting the front of the collection cylinder into the wind. Inside each trap, a 7-cm length of 1.9 cm diameter PVC pipe with a 45˚ slit cut partially through the pipe was anchored to the bottom inner surface of the trap, serving as a microscope slide holder. Slides were labeled, coated with a thin layer of silicone grease (covering approximately 75% of the length of the slide; G-697, General Electric Co., Waterford, NY, USA), and seated upright in the slide holder with the greased surface tilted back and facing upwind. The passive spore trapping device was mounted on 1.5 m (height of trap) x 1.9 cm diameter PVC pipe, which was slipped over a 1.2 m x 9.5 mm electric steel fencepost driven into the ground. With this apparatus, air moving through the trap deposited particulates, including rust urediniospores, onto the silicone grease-coated slides. Slides were collected and replaced daily at 1600 h; samples were stored in microscope slide boxes.

Active trap urediniospore collection. Model 20 Rotorod Samplers (Multidata LLC, St. Louis Park, MN 55416) consisting of a motor, a rotating sampler head, and collector rods were used to actively collect urediniospores. The motor was mounted with zip ties or hose clamps on 2 cm diameter PVC pipe slipped over metal fence posts driven into the ground, and was powered by a 12 volts battery. The leading edges of clear polystyrene collector rods (1.52 x 1.52 x 32 mm) were coated with a thin layer of silicone grease (General Electric Co.; Waterford, NY, USA).

At each of the four environments, samplers were erected at two sites (three sites for Fairhope) within the field where actively developing rust lesions were observed (Fig. A.1). Each site within a field had a pair of samplers located together with one sampler being 25% above canopy height and the other at 75% of canopy height. Within the canopy, the sampling head and rods were located above the motor and inversed for sampling over the canopy. Rotorod heads

rotated at 2400-2440 rpm. Samples were collected daily every 2 h from 08:00 to 18:00 h, except at Attapulgus, where the samples were collected at 08:30, 10:30, 12:30, 14:30, and 17:00 h. Each sampling interval lasted 30 min and the five sampling times were categorized as morning, mid-morning, noon, mid-afternoon, and late afternoon. Collector rods were subsequently adhered with white Elmer's glue (Elmer's Products, Inc., Columbus, OH, USA) to labeled microscope slides and stored in microscope slide boxes for further examination.

Urediniospore counts. For passive traps, slides were observed through a compound microscope at 200x magnification. The number of collected urediniospores were enumerated by counting 20 to 60 sample areas (each was 0.58 mm^2) depending on spore density. The represented area was 11.6 mm^2 to 34.9 mm^2 or 0.65% to 1.96% of the total 1775 mm^2 greased area. Average number of spores were expressed as urediniospores per cm^2 per day for passive traps and per m^3 per 30 min for active traps.

For active traps, urediniospores from the entire leading-edge surface of each collector rod were visually enumerated by observation under 200x magnification. Results were reported as number of spores/m^3 of air sampled per 30 min time interval. To detect differences in active trap heights and times of day, a mixed model analysis was employed (Proc Mixed SAS) on all log-transformed spore counts with trap position and time of day effects to be repeated measures. Means were separated by least significant difference (LSD) at a significance level of $\alpha = 0.05$.

Data preparation for model evaluation. The summary statistics including range and quartile of the response variables (spore counts) and all the input variables of both datasets were examined. Response variables were log scaled, adding 0.05 to the counts, to be in similar range as other variables for modeling purposes. The input variables in the passive trap dataset and active trap dataset are listed in Table 3.1. Correlations were calculated between each pair of

variables to check for multicollinearity. For model training and validation, 80% of the data points were randomly and uniformly selected from each location served as a training dataset and 20% of the data were reserved for testing model performance. Model validation was repeated 10 times, randomly partitioning the datasets 10 times, and the average prediction accuracy for different models was calculated.

LASSO model. A LASSO regression model was implemented with the GLMNET package in R v3.1.1 (R Development Core Team 2008). A "Gaussian" distribution response variable was used due to continuous output variables in the processed data. Five-fold cross validation was conducted to select the regularization parameter lambda. A LASSO regression model was fitted using the minimized lambda value to predict the spore counts of the testing dataset. After the model was fitted, the variance inflation factor (VIF) was checked for each variable that remained in the model.

Zero-inflation Poisson/regular Poisson model. This model used the "zeroinfl" function in the pscl package in R for the passive trap dataset. Only the important variables identified in the LASSO model were included to account for multicollinearity. In order to identify potential covariates that affect the probability of being in the all-zero component, analysis of variance analysis (ANOVA) was first conducted to compare the means of each predicting variable (Table A.1) in the 'always-zero' count group and 'non-zero' count group, and the variables with significant different means between the 'always-zero' count group and 'non-zero' count group were potential covariates that caused excess zeros and were included in the inflation component portion of the zero-inflated Poisson model. For the active trap dataset, a regular Poisson model was fitted using the "glm" function in R since there were few zero counts.

Random forest model. The random forest model was trained with the training dataset derived from the original data using the "caret" package in R environment. The number of variables (mtry) randomly sampled as candidates for node splitting during the tree growing process was optimized using the train function through five-fold cross validation. The numbers of trees (ntree) induced in the training process was based on a grid search ranging from 300 to 1000. Once the model was trained with the optimized number of trees and variables, spore counts in the testing dataset were predicted using the fitted model.

Neural network model. The neural network model also was developed using the "caret" package in R. Parameters considered and tuned in the neural network model were learning rate (learningrate), activation function (act.fct), number of hidden layers and number of neurons in each hidden layer (hidden), and maximum steps for the training of the neural network (stepmax). All input variables were centered and scaled by subtracting the mean from values followed by dividing values by standard deviation. The model was trained by comparing settings at three learning rates (0.1, 0.01, and 0.001) and at two stepmax (100,000 and 1,000,000) factorial combinations. The logistic activation function was used for smoothing the result of the cross product of neurons and the weights. Hidden layer and number of neurons in each hidden layer were optimized using the train function in the caret package through five-fold cross validation. The neural network model was then fitted with the optimized parameters, and spore counts in the testing datasets were predicted using the fitted model.

RESULTS

A total of 376 daily spore counts were obtained from passive traps at three environments, and 728 hourly data counts were obtained from active traps at four environments. Spore counts

ranged from 0 to 43,503 and from 0 to 19,720 for passive trap and active trap datasets, respectively. There were 161 zero counts and 8 extreme high counts (> 30,000) in the passive trap dataset, and there were 4 zero counts and 9 extreme high counts (> 10,000) in the active trap dataset. All the data points were used to build the predictive models.

Passive trap dataset. The sum of spores over both distances (15 and 61 m), all directions (eight traps for each distance), four environments, and all sampling times was over 1.0 million spores/m^2 per day with an average count per location of 42,186 spores per day. The inner ring traps had 69% of the total number of spores/m^2 while the outer ring had 31% of the total spores/m^2. Correlation between spore counts and each of the 10 input variables showed that mean daily temperature (r = -0.565; P = 0.001), mean daily relative humidity (r = -0.539; P = 0.001), and distance of trap from the source (r = -0.666; P = 0.000028) were negatively correlated with spore counts while minimum daily wind speed (r = 0.593; P = 0.02) was positively correlated with spore counts.

Active trap dataset. Spore numbers within the canopy were significantly correlated to those above the canopy in all environments in 19 of the 30 days monitored, and the spore counts within the canopy were uniformly higher than those above the canopy (Table 3.2). Temperature was positively related to spore counts on 8 of the 30 days. We observed 9 and 4 days when spore counts were significantly negatively correlated with leaf wetness and humidity, respectively. One significant correlation in opposition to general trends was detected on 12-Nov-2005 at Quincy-2, when the highest spore count was correlated with low temperature, high leaf wetness, and high humidity. There were only a few precipitation events during the sampling period, but significant correlations with spore counts were negative. Wind speed and spore count were significantly positively correlated on 9 of the 30 days.

There was considerable variation in the numbers of *P. pachyrhizi* urediniospores in the traps at the four locations. The trend lines for urediniospore counts at Fairhope and Attapulgus increased trend over time while the trend line for counts at Quincy-1 and Quincy-2 decreased (Fig. 3.1A). Along with the correlated variables temperature, humidity, leaf wetness, and precipitation (Table 3.2), mean daily spore counts in the traps were correlated with wind speed during the same time period (Fig. 3.1B).

Spore counts tended to be lower in the morning, increase through the day until afternoon or late afternoon (Fig. 3.2). Average spore counts from the active traps for each time of day in each environment and corresponding environmental variables were also correlated (Table 3.3). Seventeen of the 20 correlations between within canopy and above canopy data were positive and significant. The exceptions to this trend occurred at Quincy-1 where spore counts were low. There were no significant positive correlations of spore counts with temperature, leaf wetness, or humidity. Correlations of temperature, leaf wetness, and humidity with spore count were negative and significant for two, three, and, two occasions, respectively.

LASSO regression. The average performance of the LASSO regression model as a function of smoothing parameter lambda cross validation identified the optimized smoothing parameter lambda, which ranged from 0.05 to 0.17 with 7 to 4 variables in the model for the passive trap dataset, and ranged from 0.007 to 0.05 with 13 to 9 variables in the model for the active trap dataset (Fig. A.3). The variables that significantly contributed to the spore count at lambda = 0.05 for the passive trap dataset in decreasing order of importance were mean daily temperature, mean daily wind speed, mean daily wind direction, max and minimum wind speed, precipitation, position of trap, mean daily relative humidity, distance from source, and degree between trap and wind direction (Table 3.4). Variables that significantly contributed to active trap spore counts at

lambda = 0.007 for the active trap dataset in decreasing order of importance were precipitation, height of trap, time of day, mean previous two-hour temperature, previous second-hour temperature, mean hourly wind speed, mean hourly temperature, mean hourly relative humidity, previous hour temperature, leaf wetness, mean hourly wind direction, solar radiation, and previous hour humidity (Table 3.4).

Selected variables used to fit regression model and predict spore counts in the testing dataset explained 48.2% and 47.3% of the variation for the passive dataset and active dataset, respectively (Fig. 3.3). The prediction accuracy of the LASSO regression model evaluated by 10 repeats of validation ranged from 0. 28 to 0.71 for the passive trap dataset, and 0.61 to 0.70 for the active trap dataset (Table 3.5). The average prediction accuracy was 0.56 and 0. 65 for the passive and active trap dataset, respectively.

Zero-inflation Poisson/regular Poisson model. Daily maximum wind speed and temperature were the only two variables that were significantly different between 'always-zero' group and the 'non-zero' group (Table A.1), therefore, these two variables were included in the inflation component portion. All the predictors included in the Poisson count portion of the zero-inflated Poisson model were statistically significant for the passive trap dataset ($P < 0.0001$), and daily maximum wind speed in the inflation component portion was also significant ($P = 0.0323$) (Table A.2). The estimated odds of 'excess zero' was about exp (-0.11) = 0.89 times lower when the wind speed increased by one unit. All the parameters included in the Poisson model for the active trap dataset were significant ($P < 0.0001$). The prediction accuracy of the zero-inflated Poisson model evaluated by 10 repeats of validation ranged from 0.73 to 0.83 for the passive trap dataset, and the average prediction accuracy was 0.77. The prediction accuracy of the regular Poisson model evaluated by 10 repeats of validation ranged from and 0.69 to 0.82 for the active

trap dataset, and 0. 74 for the passive and active trap dataset (Table 3.5). The predicted values of the 10 testing datasets by the zero-inflation Poisson /regular Poisson model were plotted against the actual spore count values for the passive trap and active trap datasets (Fig. 3.4).

Random forest. The optimized number of variables used at each node split in the final random forest model was m = 6 and m = 14 for the passive and active trap datasets, respectively (Fig. A.4), and 400 trees minimized the MSE of the random forest model for both datasets. The final model explained 76.1% and 86.8% of the total variation for passive and active trap datasets, respectively. The predicted values of the 10 testing datasets by the random forest model were plotted against the actual spore count values for the passive trap and active trap datasets (Fig. 3.5). The prediction accuracy of the random forest model evaluated by 10 repeats of validation ranged from 0.75 to 0.86 for the passive trap dataset, and 0.93 to 0.95 for the active trap dataset (Table 3.5). The average prediction accuracy was 0.80 and 0. 94 for the passive and active trap dataset, respectively. In decreasing order of importance, the variables contributing to the model were mean daily temperature, minimum wind speed, mean daily wind direction, maximum and minimum wind speed, precipitation, position of trap, mean daily relative humidity, distance from source, and degree between trap and wind (Table 3.4). For the active trap dataset, variables in decreasing order of importance were mean hourly relative humidity, previous two-hour mean temperature, mean hourly temperature, mean hourly wind direction, previous second hour mean temperature, previous hour humidity, solar radiation, wind speed, previous hourly mean temperature, mean previous two-hour humidity, time of day, previous second-hour humidity, height of trap, hourly mean leaf wetness, and hourly total precipitation (Table 3.4). Like the LASSO regression model, the most important factors that affected the amount of rust spores in

the air were temperature, previous hour temperature, wind speed, wind direction, humidity, and time of day.

Neural network. The optimal parameters for the neural network model selected by 5-fold cross validation was with one hidden layer with 14 and 19 neurons in the hidden layer for the passive and active trap datasets, respectively (Fig. A.5). The model converged relatively fast with learning rate = 0.001 and maximum step of stopping learning = 1,000,000. The predicted values of the 10 testing datasets by the neural network model were plotted against the actual spore count values for the passive trap and active trap datasets (Fig. 3.6). The prediction accuracy of the neural network model evaluated by 10 repeats of validation ranged from 0.76 to 0.83 for the passive trap dataset, and 0.57 to 0.71 for the active trap dataset (Table 3.5). The average prediction accuracy was 0.79 and 0. 63 for the passive and active trap dataset, respectively. The neural network model indicated that the most important predictor variables that affected spore counts in the passive traps in decreasing order were mean daily temperature, minimum wind speed, mean daily wind speed, distance from source, mean daily relative humidity, precipitation, maximum wind speed, mean daily wind direction, degree between trap and wind, and position of trap (Table 3.4). The variables that significantly contributed to the spore count for the passive trap dataset in decreasing order of importance were mean daily temperature, mean daily relative humidity, distance from source, minimum wind speed, maximum wind speed, mean daily wind direction, position of the trap, precipitation, mean daily wind speed, degree between trap, and wind direction (Table 3.4). Variables that significantly contributed to active trap spore counts for the active trap dataset were mean hourly relative humidity, mean previous two-hour temperature, previous hour humidity, mean hourly wind direction, previous second hour temperature, mean previous two-hour humidity, mean hourly

temperature, previous hour temperature, previous second hour humidity, height of trap, solar radiation, mean hourly wind speed, time of day, and leaf wetness (Table 3.4).

DISCUSSION

The short-distance movement of the *P. phychyrhizi* urediniospores from infected soybean leaves to the surrounding air determines the inoculum availability for short- and long-distance spore dispersion and is dependent on many factors. Our results showed that temperature, wind speed, wind direction, relative humidity, leaf wetness, time of day, and distance from the source were the most important factors that affected urediniospore movement. There were positive correlations between the number of urediniospore with temperature and wind speed, and negative correlations between number of urediniospore with humidity, leaf wetness, and precipitation. These factors in combination resulted in a greater majority of the urediniospores counted between midday to mid-afternoon. Along with these factors, there may be additional biological explanations for higher spore counts around midday. In other studies, it was shown that *Puccinia striiformis* produces a mucilaginous layer around its spores that becomes thicker as relative humidity increases (Rapilly 1979), which may also be the case for *P. pachyrhizi* (Del Ponte et al. 2006b), resulting in fewer numbers of spores released into the air.

It is not uncommon for fungi to have a diurnal pattern of peak spore release. Similar to what was found in our study, this was shown for capture of *Podosphaera clandestina* (powdery mildew of sweet cherry) conidia, where spore numbers peaked about midday, which was positively correlated with wind speed and temperature, but negatively correlated to relative humidity (Grove 1998). In another example, a diurnal pattern was shown for peanut rust as spore capture peaked at noon, which was positively correlated to relative humidity below 70% and to

increased wind speed (Savary and Janeau 1986). In the case of *P. pachyrhizi*, newly formed urediniospores push older spores upward often forming a chain under still conditions. When the chain is disturbed, the urediniospores become airborne. As the spore chain gets longer, more urediniospores are released into the air.

Precipitation events are generally perceived to wash spores from the atmosphere down to surfaces, including plant leaves. One report showed that precipitation washed *P. pachyrhizi* urediniospores from the leaves and out of the atmosphere (Default et al. 2010), but several reports speculate that rain events may disperse *P. pachyrhizi* urediniospores by rain splash (Del Ponte and Esker 2008; Melching et al. 1979). In our study, there were only a few precipitation events and trends related to precipitation were not easily discerned, although when precipitation occurred, it was associated with a reduction in urediniospore counts. However, precipitation and high moisture conditions are important in soybean rust epidemics as shown in several soybean field studies with *P. pachyrhizi* in Brazil (Del Ponte et al. 2006b), China (Tan et al. 1996), Taiwan (Tschanz et al. 1984), and Uganda (Kawuki et al. 2004), where precipitation events resulted in higher rust severity after the event. Another key role played by precipitation is washing urediniospores from the air and onto host leaves, and providing the leaf wetness necessary for urediniospore germination.

Disease forecasting models provide a way to predict the risk of disease. Regression models have been widely used to predict epidemics of plant diseases including soybean rust (Del Ponte et al. 2006a; Olatinwo et al. 2008; Paul and Munkvold 2005; Rosso and Hansen 2003; Uddin et al. 2003). Regression models have the advantage of simplicity and easy interpretation of predicted results. However, when response and predicting variables have relationships that are more complex than linear, other approaches such as random forest and neural network have been

used in making more accurate predictions in many areas, including plant disease epidemiology (Behmann et al. 2015; Mehra et al. 2016; Paul and Munkvold 2005). In our datasets, many significant interactions were detected between variables. The higher prediction accuracy of the random forest and neural network models over the linear model for predicting rust urediniospore counts in the traps shows the power of the non-linear models to help model these complex interactions.

Our study developed and compared four machine learning models for the prediction of short distance movement of soybean rust urediniospores. The random forest model produced more accurate predictions than the LASSO regression model, the zero-inflated Poisson model, and the neural network model indicating that the relationships between the number of urediniospores released from the leaf surface into the air and the various meteorological variables are more complex than simply linear, and demonstrated the robustness of the random forest model to handle multicollinearity and inflated-zero issues. These four models all identified temperature, humidity, wind speed, wind direction, and leaf wetness as key factors affecting the rust urediniospores collected by the passive and active spore traps. The random forest and neural network models provide a way to analyze non-linear relationships between variables (Olden et al. 2008; Siroky 2009). However, it can be difficult to interpret the relationships between variables identified by these models due to the lack of simple representation, such as a linear formula. In addition, the values of many parameters in both random forest and neural network models are highly influential. The number of trees and number of variables used at each node in the random forest model and the learning rate, activation function, number of hidden layers, and number of neurons in each layer can greatly affect prediction accuracy.

Overall, we conclude that using multiple machine learning techniques was helpful, not only to identify the most important variables and make the most accurate predictions, but also to increase confidence for applying these models to obtain coherent results. The models built in our study are useful in predicting the numbers of soybean rust urediniospores found a short distance from the source, which in turn is critical for understanding long distance dispersion of the spores including for prediction and forecasting models that are used to advise producers.

ACKNOWLEDGEMENTS

We thank L. Beck, M. R. Miles, and T. Steinlage for their assistance in data collection and the personnel at the university experiment stations at Fairhope, Alabama; Quincy, Florida; and Attapulgus, Georgia. We would also like to thank T. Herman for edits and comments on this manuscript.

TABLES

Table 3.1. Input variables in the datasets for each model collected from the passive trap and active trap.

Dataset	Input Variables
Passive Trap	1. Daily mean temperature (˚C) 2. Daily mean relative humidity (%) 3. Daily mean precipitation (mm) 4. Daily mean wind speed (km/h) 5. Daily mean wind direction (˚) 6. Minimum wind speed (km/h) 7. Maximum wind speed (km/h) 8. Position of trap (˚) 9. Distance (m) 10. Difference between trap location and wind (˚)
Active Trap	1. Time of day 2. Hourly mean temperature (˚C) 3. Hourly mean relative humidity (%) 4. Hourly total precipitation (mm) 5. Hourly mean wind speed (km/h) 6. Hourly mean wind direction (˚) 7. Trap height (m) 8. Hourly mean leaf wetness (Centibar) 9. Hourly mean solar radiation (Wat/m^2) 10. Previous hour mean temperature (˚C) 11. Previous hour mean relative humidity (%) 12. Previous second hour mean temperature (˚C) 13. Previous second hour mean relative humidity (%) 14. Previous two hours mean temperature (˚C) 15. Previous two hours mean relative humidity (%)

Table 3.2. Active trap daily mean *Phakopsora pachyrhizi* urediniospores concentrations at both canopy heights and all environments, correlated to within canopy spore concentration, temperature, leaf wetness, humidity, precipitation, and wind speed.

		Canopy	Mean	Correlation					
Environment	**Date**	**Position**	**Spores/m³**	**Within Canopy**	**Temperature**	**Leaf Wetness**	**Humidity**	**Rain**	**Wind Speed**
Fairhope	11-Aug-05	Above	94.88	0.26	0.78	-0.46	-0.7	-0.34	0.58
		Within	181.94		-0.33	0.74	0.4	0.81	-0.5
	12-Aug-05	Above	100.69	0.87	0.39	-0.62	-0.48		0.69
		Within	347.41		-0.09	-0.18	-0.03		0.27
	13-Aug-05	Above	178.12	0.99**	0.88*	-0.98**	-0.78		0.8
		Within	379.51		0.79	-0.94*	-0.67		0.79
	15-Aug-05	Above	204.91	0.86	0.27	0.34	-0.76	0.46	0.98**
		Within	549		0.18	0.4	-0.53	0.48	0.79
	16-Aug-05	Above	11.06	0.36	0.64	-0.63	-0.72	0.65	0.75
		Within	30.77		0.34	-0.26	-0.34	-0.46	0.78
	17-Aug-05	Above	58.11	0.93*	-0.25	-0.29	-0.05	-0.01	-0.81
		Within	73.96		-0.15	-0.22	-0.14	0	-0.66
	18-Aug-05	Above	50.28	0.98**	0.61	-0.62	-0.48		0.67
		Within	64.35		0.65	-0.63	-0.49		0.66
	19-Aug-05	Above	22.34	0.92*	0.92*	-0.67	-0.93*	-0.09	0.45
		Within	23.34		0.78	-0.72	-0.82	-0.29	0.34
Attapulgus	5-Oct-05	Above	303.26	0.93*	-0.01	0.39	0.12		0.23
		Within	811.85		0	0.61	0.16		0.38
	6-Oct-05	Above	38.9	0.99**	0.79	-0.82	-0.77	-0.93*	0.86
		Within	118.93		0.87	-0.9*	-0.85	-0.9*	0.84
	7-Oct-05	Above	16.56	0.96*	0.47	-0.71	-0.4	-0.4	0.76
		Within	51.32		0.65	-0.64	-0.6	-0.26	0.73
	8-Oct-05	Above	32.9	0.7	0.58	-0.72	-0.56		0.64
		Within	66.18		0.53	-0.24	-0.54		0.49
	9-Oct-05	Above	36.95	1**	0.83	-0.64	-0.8		0.17
		Within	126.43		0.85	-0.65	-0.83		0.2
	11-Oct-05	Above	52.35	0.98**	0.32	-0.73	-0.21		0.45
		Within	142.05		0.22	-0.69	-0.11		0.42
Quincy-1	5-Oct-05	Above	23.51	0.9*	-0.1	0.79	0.12	0.81	0.19
		Within	30.07		-0.17	0.78	0.09	0.79	0.12
	6-Oct-05	Above	17.22	0.98**	0.09	-0.97**	-0.96**		0.73
		Within	25.66		0.22	-0.95*	-0.94*		0.8
	7-Oct-05	Above	3.79	0.59	0.81	-0.58	-0.9*		-0.59
		Within	7.76		0.12	0.24	-0.49		-0.18
	8-Oct-05	Above	7.98	0.8	0.53		0.14		0.07
		Within	25.76		0.35		0.36		0.06
	11-Oct-05	Above	25.29	0.96*	0.93*	-0.91*	-0.87		0.96*
		Within	93.22		0.82	-0.78	-0.77		0.86
	12-Oct-05	Above	25.65	0.77	0.21	-0.08	-0.24		0.73
		Within	97.65		0.05	-0.26	0.02		0.69
Quincy-2	2-Nov-05	Above	310.12	0.89*	-0.35		0.54		0.93
		Within	1106.68		0.29		-0.12		0.5
	3-Nov-05	Above	821.92	0.93*	-0.43	0.07	0.52		0.89*
		Within	2228.37		-0.57	0.23	0.64		0.87
	4-Nov-05	Above	557.87	0.72	0.89*	-0.64	-0.86		0.85
		Within	2961.87		0.77	-0.97**	-0.83		0.33
	6-Nov-05	Above	1688.61	0.85	0.32	-0.33	-0.02		0.97**
		Within	6277.78		0.66	-0.69	-0.3		0.86
	7-Nov-05	Above	2225.24	0.99**	0.97**	-0.92*	-0.87		0.41
		Within	8085.5		0.97**	-0.89*	-0.84		0.32
	8-Nov-05	Above	1256.53	0.98*	0.67	-0.8	-0.51		0.99*
		Within	5335.04		0.54	-0.73	-0.36		1**
	9-Nov-05	Above	1437.29	0.99**	0.81	-0.76	-0.76		0.62
		Within	3490.65		0.8	-0.78	-0.73		0.56
	10-Nov-05	Above	9446.19	0.98**	0.89*	-0.83	-0.73		0.89*
		Within	16138.47		0.9*	-0.86	-0.75		0.96**
	11-Nov-05	Above	4125.74	0.92	-0.88		0.83		0.98*
		Within	5956.31		-0.85		0.91		0.89
	12-Nov-05	Above	1993.88	0.96*	-0.86	0.82	0.85		-0.31
		Within	4056.81		-0.96*	0.92*	0.94*		-0.57

Table 3.3. Time of day mean spore concentrations in the active traps at both canopy heights and all environments, correlated to within canopy spore concentration, temperature, leaf wetness, humidity, precipitation, and wind speed

Environment	Time of Day[a]	Canopy Position	Mean Spores/m^3	Correlations						
				Within Canopy	Temperature	Leaf Wetness	Humidity	Rain	Wind Speed	
Fairhope	Morning	Above	24.53	0.93**	0.1	-0.02	-0.22	0.39	-0.07	
		Within	75.22		-0.16	0.24	0.01	0.31	-0.37	
	MidMorning	Above	61.69	0.92**	0.18	-0.01	-0.12		0.52	
		Within	118.54		0.09	-0.16	-0.17		0.25	
	Noon	Above	71.32	0.95**	0.06		0.12	-0.43	-0.02	
		Within	120.56		0.25		-0.03	-0.57	0.2	
	Afternoon	Above	139.86	0.93**	0.5	-0.8*	-0.56		0.38	
		Within	226.28		0.24	-0.59	-0.32		0.35	
	LateAfternoon	Above	68.68	0.93**	-0.2	0	0.05	0.45	0.77	*
		Within	106.8		-0.23	0.24	0.04	0.51	0.61	
Attapulgus	Morning	Above	13.02	0.93**	0.03	-0.69	-0.86*	-0.42	0.81	
		Within	62.18		-0.16	-0.57	-0.78	-0.19	0.75	
	MidMorning	Above	66.38	0.97**	0.68	-0.22	-0.41		0.79	
		Within	138.52		0.75	-0.22	-0.58		0.69	
	Noon	Above	119.14	0.89*	0.13		-0.43		0.86	*
		Within	282.5		0.29		-0.4		0.86	*
	Afternoon	Above	59	0.94**	-0.04	-0.96**	-0.49	-0.96	** 0.85	*
		Within	147.86		-0.11	-0.87*	-0.38	-0.87	* 0.96	**
	LateAfternoon	Above	39.37	1**	-0.8	0	0.71		0.47	
		Within	135.47		-0.84*	-0.05	0.69		0.5	
Quincy-1	Morning	Above	4.21	0.69	0	-0.63	-0.37		0.61	
		Within	12.59		-0.56	-0.36	0.37		0.02	
	MidMorning	Above	14.2	0.84*	-0.67	-0.47	-0.07		0.6	
		Within	32.27		-0.82*	-0.16	0.23		0.13	
	Noon	Above	16.51	0.77	0.37	0.65	-0.68		0.43	
		Within	41.48		-0.01	0.28	-0.76		0.35	
	Afternoon	Above	30.72	0.89*	-0.12	0.16	-0.58		0.43	
		Within	67.02		-0.4	-0.19	-0.88*		0.49	
	LateAfternoon	Above	39.02	0.67	0.08	0.44	0.49	0.39	0.04	
		Within	53.25		-0.34	0.68	0.23	0.67	0.75	
Quincy-2	Morning	Above	592.91	0.84**	-0.18	-0.03	-0.03		0.74	*
		Within	2250.9		-0.26	0.22	0.34		0.59	
	MidMorning	Above	2684.59	0.92**	-0.32	-0.17	0.07		0.41	
		Within	5592.45		-0.19	-0.09	0.31		0.11	
	Noon	Above	2048.23	0.96**	0.17	0.31	0.47		-0.03	
		Within	4841.73		0.26	0.4	0.52		-0.18	
	Afternoon	Above	1625.6	0.95**	0.11		0.33		0.62	
		Within	4462.24		0.15		0.54		0.63	
	LateAfternoon	Above	802.88	0.75*	-0.03	0.25	-0.1		0.54	
		Within	3144.28		-0.07	0.41	0.44		0.19	

[a]Morning=at 8:00 am to 8:30 am
Mid-morning=at 10:00 am to 10:30am
Noon = 12:00 pm to 12:30 pm
Afternoon 14:00 pm to 14:30 pm
Late Afternoon 16:00 pm to 16:30 pm

Table 3.4. Relative importance of variables for predicting the spore counts in the spore traps as identified by the lasso regression, random forest, and neural network models. Variables importance was determined using the "varImp" function in the caret package in R, and all the measures of importance are scaled to have a maximum value of 100.

	Lasso Regression		Random forest		Neural network	
Dataset	Variables	Relative importance (%)	Variables	Relative importance (%)	Variables	Relative importance (%)
Passive trap	Mean daily wind speed	100.0	Mean daily temperature	100.0	Mean daily temperature	100.0
	Mean daily temperature	56.0	Mean daily wind Speed	67.5	Min wind speed	48.4
	Mean daily wind direction	18.6	Mean daily wind direction	34.8	Mean daily wind speed	46.3
	Distance from source	15.5	Max wind speed	32.1	Distance from source	38.2
	Min wind speed	13.3	Min wind speed	28.8	Mean daily relative humidity	35.8
	Mean daily relative humidity	6.0	Precipitation	19.7	Precipitation	6.9
	Max wind speed	5.3	Position of trap	14.9	Max wind speed	6.4
	Position of trap	0.0	Mean daily relative humidity	13.58	Mean daily weed direction	3.0
	Precipitation	0.0	Distance from source	13.23	Degree between trap and wind	2.6
	Degree between trap and wind	0.0	Degree between trap and wind	4.2	Position of trap	0.0
Active trap	Precipitation	100.0	Mean hourly relative humidity	100.0	Mean hourly relative humidity	100.0
	Height of trap	27.5	Mean previous two hour temperature	69.7	Mean previous two hour temperature	92.8
	Time of day	3.9	Mean hourly temperature	54.8	Previous hour humidity	78.6
	Mean previous two hour temperature	3.8	Mean hourly wind direction	48.5	Mean hourly wind direction	71.4
	Previous second hour temperature	3.6	Previous second hour temperature	43.0	Previous second hour temperature	68.7
	Mean hourly wind speed	2.6	Previous hour humidity	38.3	Mean previous two hour humidity	68.4
	Mean hourly temperature	2.6	Solar radiation	36.8	Mean hourly temperature	62.8
	Mean hourly relative humidity	1.6	Mean hourly wind speed	31.8	Previous hour temperature	55.5
	Previous hour temperature	1.4	Previous hour temperature	30.4	Previous second hour humidity	52.3
	Leaf wetness	1.3	Mean previous two hour humidity	27.4	Height of trap	45.2
	Mean hourly wind direction	1.1	Time of day	25.9	Solar radiation	38.5
	Solar radiation	0.9	Previous second hour humidity	20.1	Mean hourly wind speed	11.7
	Previous hour humidity	0.5	Height of trap	15.7	Time of day	4.3
	Previous second hour humidity	0.0	Leaf wetness	11.7	Leaf wetness	2.5
	Previous two hour humidity	0.0	Precipitation	1.4	Precipitation	0.0

Table 3.5. Prediction accuracy of the LASSO, zero-inflated Poisson/regular Poisson model, random forest, and neural network models in 10 validation process for the passive trap dataset and active trap dataset. In each validation, 80% of the data points were randomly and uniformly selected from each location and used as training dataset; 20% of the data was saved for testing model performance. Prediction accuracy was represented as the correlation coefficients between the predicted spore counts and the actual spore counts.

	Passive Trap Dataset				Active Trap Dataset			
	LASSO	Zero-inflated Poisson	Random Forest	Neural Network	LASSO	Poisson	Random Forest	Neural Network
1	0.5	0.75	0.78	0.81	0.28	0.76	0.94	0.67
2	0.28	0.77	0.83	0.77	0.5	0.76	0.93	0.57
3	0.59	0.77	0.85	0.77	0.59	0.73	0.93	0.6
4	0.42	0.79	0.88	0.76	0.42	0.75	0.94	0.66
5	0.53	0.79	0.78	0.78	0.53	0.83	0.94	0.58
6	0.72	0.83	0.91	0.77	0.72	0.7	0.94	0.58
7	0.72	0.74	0.82	0.76	0.72	0.72	0.94	0.66
8	0.63	0.74	0.82	0.83	0.62	0.7	0.95	0.71
9	0.65	0.79	0.81	0.79	0.65	0.73	0.94	0.68
10	0.65	0.73	0.82	0.83	0.53	0.69	0.93	0.6
Avg.	0.56	0.77	0.83	0.79	0.65	0.74	0.94	0.63

FIGURES

Fig. 3.1. Enumeration of *Phakopsora pachyrhizi* urediniospores and wind speed over time in four environments. **A,** Mean daily urediniospore counts in active rotorod (above and below soybean canopy level) traps. **B,** Mean daily wind speed.

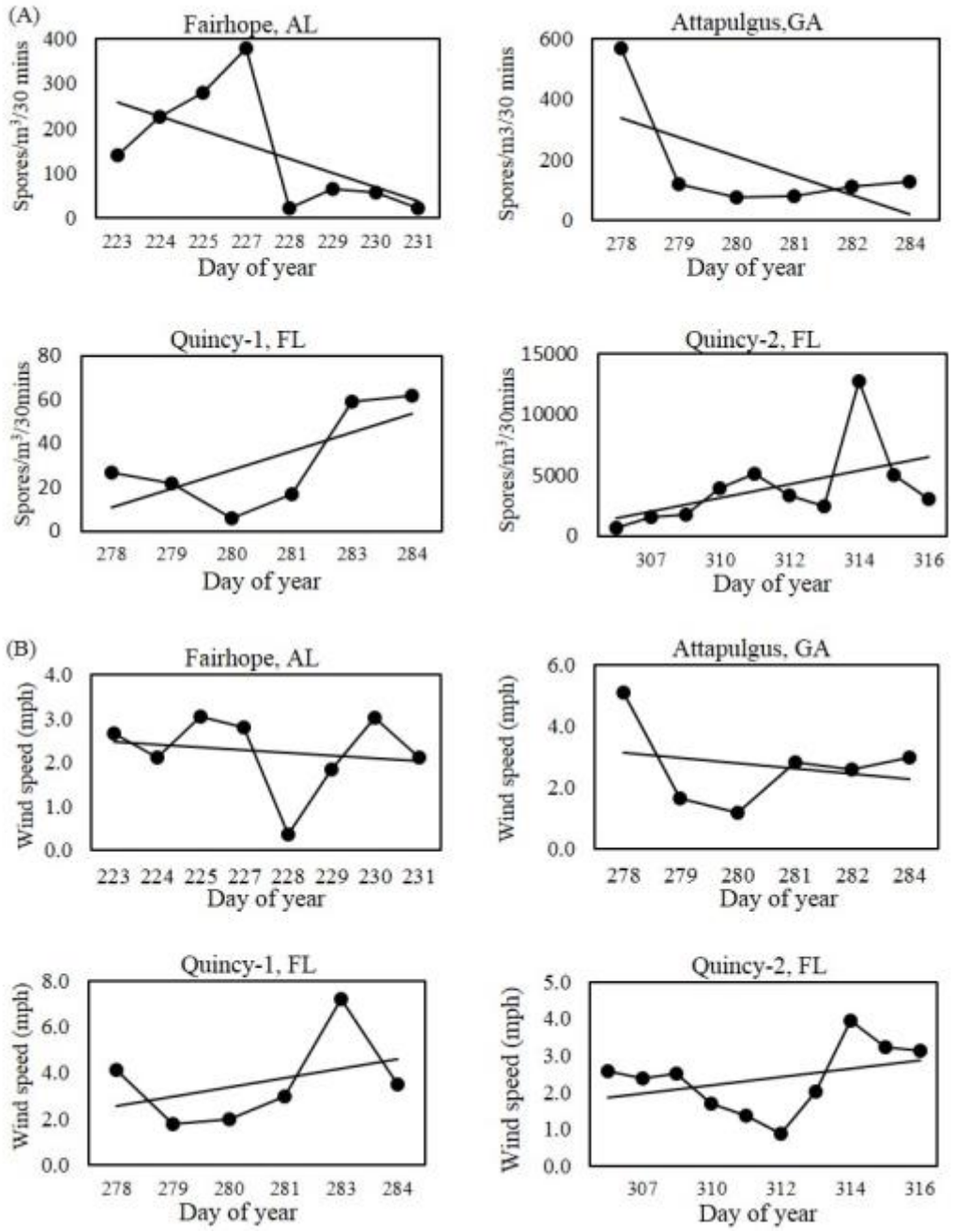


Fig. 3.2. Mean *Phakopsora pachyrhizi* urediniospores within canopy (black bars) and above canopy (gray bars) during the day at four environments. Same lowercase letters above bars within a time of day category indicate same mean spore counts ($\alpha = 0.05$) within and above the canopy. Same uppercase letters among time of day categories indicate same mean spore counts ($\alpha = 0.05$) over the categories.

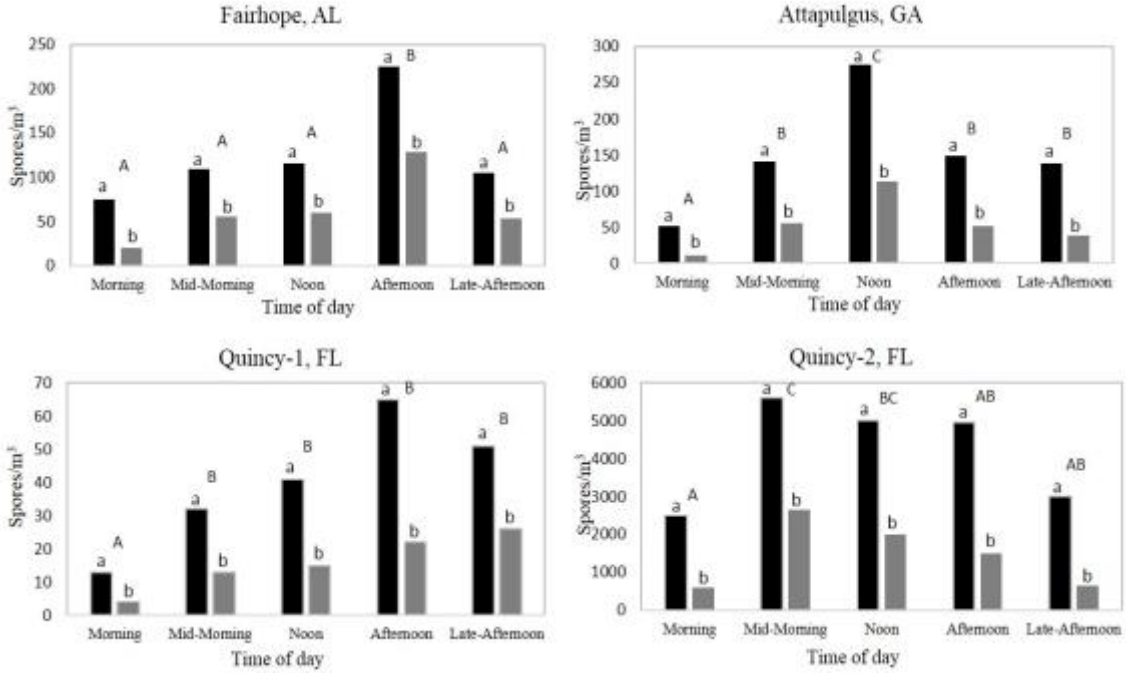


Fig. 3.3. *Phakopsora pachyrhizi* urediniospore counts predicted by the LASSO regression model vs. actual spore counts. **A,** The actual spore counts and the predicted spore counts in the plot were a combination of the actual and predicted values generated in the ten repeated model validation process for the passive trap dataset. **B,** The actual spore counts and the predicted spore counts in the plot were a combination of the actual and predicted values generated in the ten repeated model validation process for the active trap dataset.

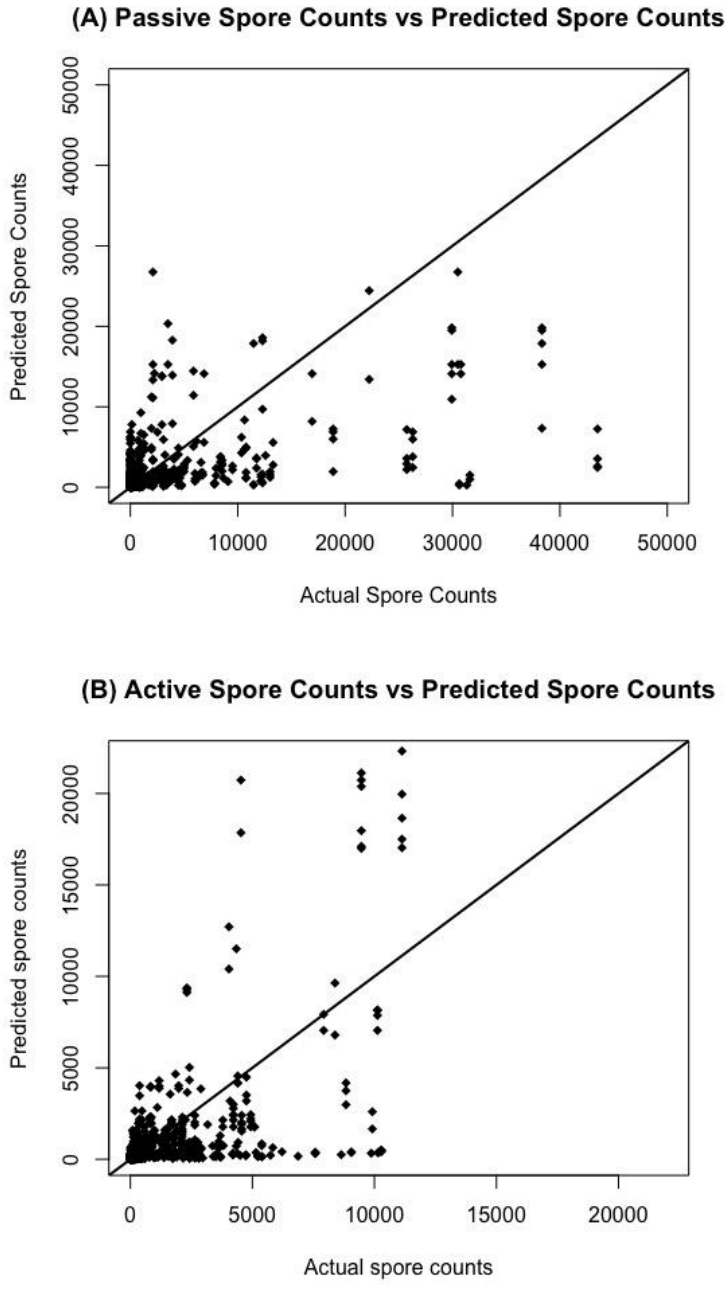
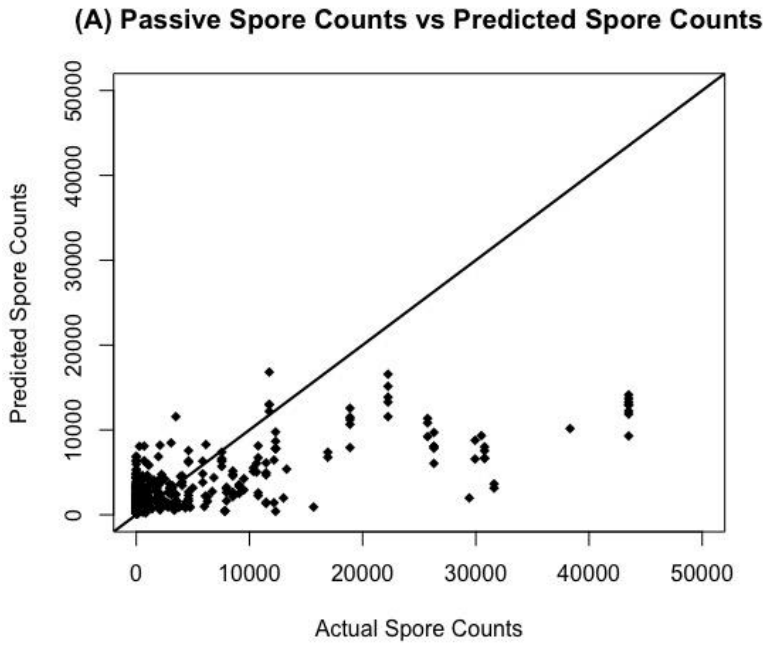


Fig. 3.4. *Phakopsora pachyrhizi* urediniospore counts predicted by the zero-inflation Poisson /regular Poisson model vs. actual spore counts. A, The actual spore counts and the predicted spore counts in the plot were a combination of the actual and predicted values generated in the ten repeated zero-inflation Poisson model validation process for the passive trap dataset. B, The actual spore counts and the predicted spore counts in the plot were a combination of the actual and predicted values generated in the ten repeated regular Poisson model validation process for the active trap dataset.



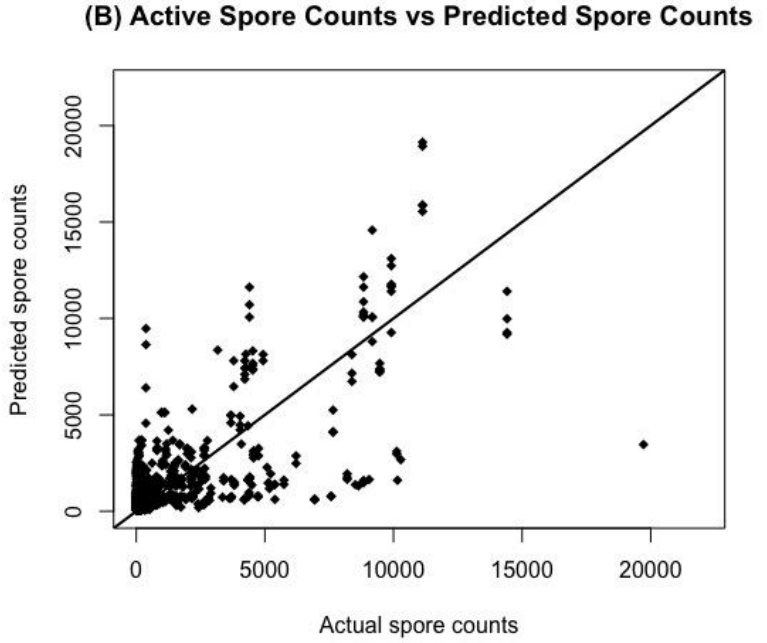
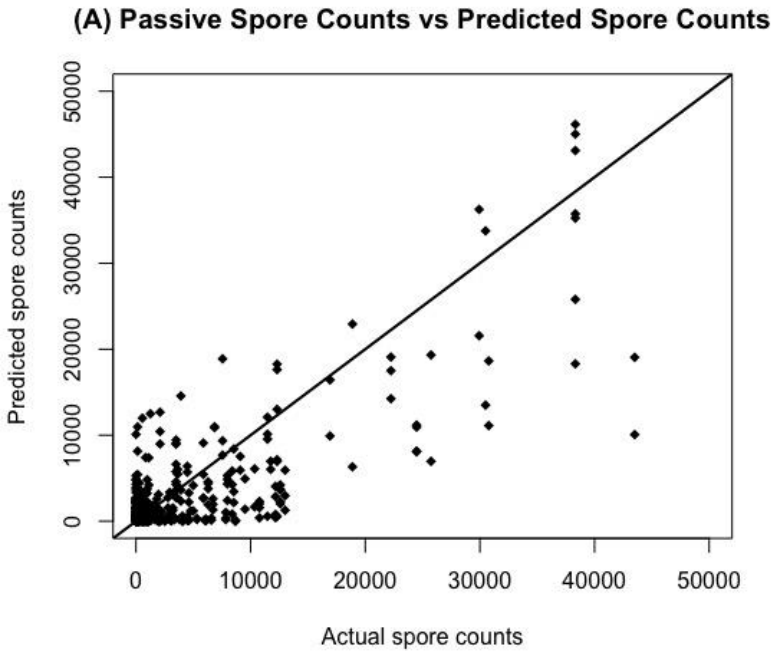


Fig. 3.5. *Phakopsora pachyrhizi* urediniospore counts predicted by the random forest model vs. actual spore counts. **A,** The actual spore counts and the predicted spore counts in the plot were a combination of the actual and predicted values generated in the ten repeated model validation process for the passive trap dataset. **B,** The actual spore counts and the predicted spore counts in the plot were a combination of the actual and predicted values generated in the ten repeated model validation process for the active trap dataset.



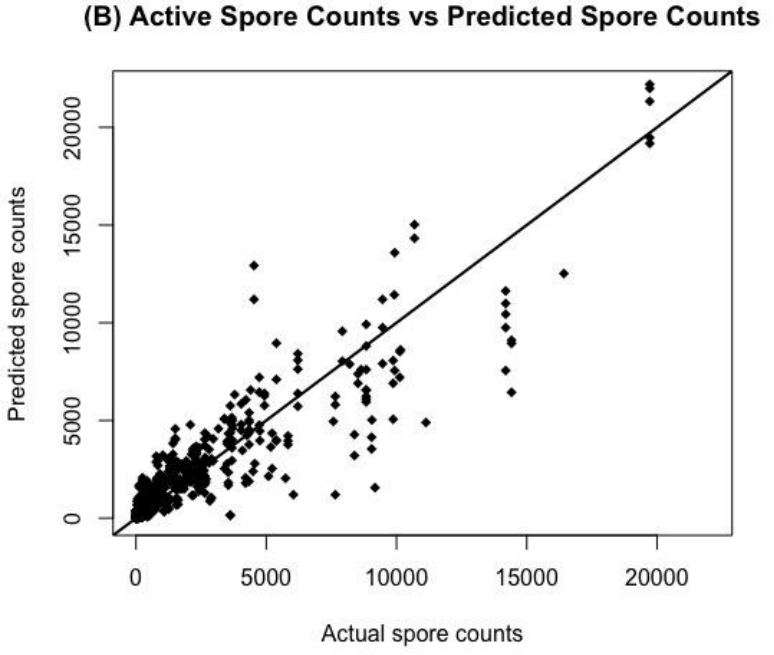
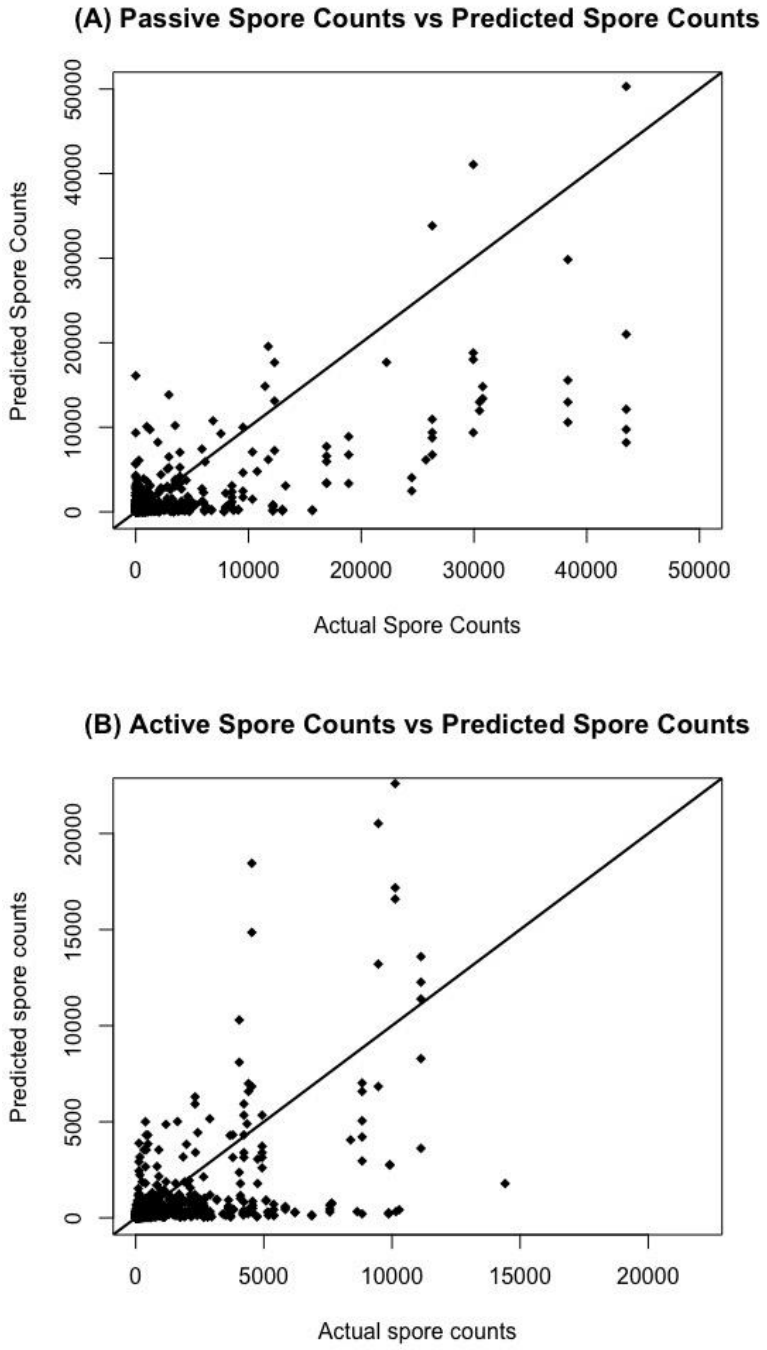


Fig. 3.6. *Phakopsora pachyrhizi* urediniospore counts predicted by the neural network model vs. actual spore counts. A, The actual spore counts and the predicted spore counts in the plot were a combination of the actual and predicted values generated in the ten repeated model validation process for the passive trap dataset. B, The actual spore counts and the predicted spore counts in the plot were a combination of the actual and predicted values generated in the ten repeated model validation process for the active trap dataset.



REFERENCES

Andrade, D., Pan, Z., Dannevik, W., and Zidek, J. 2009. Modeling soybean rust spore escape from infected canopies: model description and results. J. Appl. Meteor. Climatol. 48:789-803.

Aylor, D. E. 1986. A framework for examining inter-regional aerial transport of fungal spores. Agr. Forest Meteorol. 38:263-288.

Aylor, D. E. 1999. Biophysical scaling and the passive dispersal of fungus spores: relationship to integrated pest management strategies. Agr. Forest Meteorol. 97:275-292.

Behmann, J., Mahlein, A. K., Rumpf, T., Römer, C., and Plümer, L. 2015. A review of advanced machine learning methods for the detection of biotic stress in precision crop protection. Precis. Agric. 16:239-260.

Breiman L. 2001. Random forests. Machine Learning 45:5–32.

Christiano, R., and Scherm, H. 2007. Quantitative aspects of the spread of the Asian soybean rust in the southeastern United States, 2005 to 2006. Phytopathology 97:1428-1433.

Del Ponte, E. M., and Esker, P. D. 2008. Meteorological factors and Asian soybean rust epidemics - a system approach and implications for risk assessment. Sci. Agric. 65:88-97.

Del Ponte, E. M., Godoy, C. V., Canteri, M. G., Reis, E. M., and Yang, X. B. 2006a. Models and applications for risk assessment and prediction of Asian soybean rust epidemics. Fitopatol. Bras. 31:533-544.

Del Ponte, E. M., Godoy, C. V., Li, X., and Yang, X. B. 2006b. Predicting severity of Asian soybean rust with empirical rainfall models. Phytopathology 96:797-803.

Desborough, P. J. 1984. Selection of soybean cultivar and sowing date as a strategy for avoidance of rust (*Phakopsora pachyrhizi* Syd.) losses in coastal New South Wales. Aust. J. Exp. Agric. 24:433-439.

Dufault, N., Isard, S., Marois, J., and Wright, D. 2010. Removal of wet deposited *Phakopsora pachyrhizi* urediniospores from soybean leaves by subsequent rainfall. Plant Dis. 94:1336-1340.

Garzón, M. B., Blazek, R., Neteler, M., de Dios, R. S., Ollero, H. S., and Furlanello, C. 2006. Predicting habitat suitability with machine learning models: the potential area of *Pinus sylvestris* L. in the Iberian Peninsula. Ecol. Model. 197:383–393.

Gregory, P. H. 1973. The microbiology of the atmosphere. New York: Wiley. 377p.

Grove, G. G. 1998. Meteorological factors affecting airborne conidia concentrations and the latent period of *Podosphaera clandistina* on sweet cherry. Plant Dis. 82:741-746.

Hartman, G. L., Sikora, E. F., and Rupe, J. C. 2015. Rust. Pages 56-59 in: Compendium of Soybean Diseases and Pests, G. L. Hartman, E. F. Sikora and J. C. Rupe, eds. American Phytopathological Society, St. Paul.

Hartman, G. L., Wang T. C., and Tschanz A. T. 1991. Soybean rust development and the quantitative relationship between rust severity and soybean yield. Plant Dis. 75:596-600.

Isard, S.A., Barnes, C. W., Hambleton, S., Ariatti, A., Russo, J. M., Tenuta, A., Gay, D. A., and Szabo. L. J., 2011. Predicting seasonal soybean rust incursions into the North American continental interior using sentinel plot monitoring, spore trapping, and aerobiological modeling. Plant Dis. 95:1346-1357.

Isard, S. A., Dufault, N. S., Miles, M. R., Hartman, G. L., Russo, J. M., De Wolf, E. D., and Morel, W. 2006a. The effect of solar irradiance on the mortality of *Phakopsora pachyrhizi* urediniospores. Plant Dis. 90:941-945.

Isard, S. A. and Gage, S. H. 2001. Flow of life in the atmosphere: An airscape approach to understanding invasive organisms. Michigan State University Press, East Lansing, MI.
Isard, S. A., Gage, S. H., Comtois, P., and Russo, J. M. 2005. Principles of the atmospheric pathway for invasive species applied to soybean rust. BioSci. 55:851-861.
Isard, S. A., Russo, J. M., and Ariatti, A. 2007. The Integrated Aerobiology Modeling System applied to the spread of soybean rust into the Ohio River valley during September 2006. Aerobiologia 23:271-282.
Isard, S. A., Russo, J. M., and DeWolf, E. D. 2006b. The establishment of a national pest information platform for extension and education. Online, Plant Health Progress, doi:10.1094/PHP-2006-0915-01-RV.
Kalnay, E. 1996. The NCEP/NCAR 40-year reanalysis project. B. Am. Meteorol. Soc. 82:247-279.
Kanamitsu, M., Ebisuzaki, W., Wollen, J., Yang, S-K., Hnilo, J.J., Fiorino, M., Potter, G. L. 2002. NCEP-DOE AMIP-II Reanalysis (R-2). B. Am. Meteorol. Soc. 83:1631-1643.
Kawuki, R. S., Tukamuhabwa, P., and Adipala, E. 2004. Soybean rust severity, rate of rust development, and tolerance as influenced by maturity period and season. Crop Prot. 23:447-455.
Kelly, H. Y., Dufault, N. S., Walker, D. R., Isard, S. A., Schneider, R. W., Giesler, L. J., Wright, D. L., Marios, J. J., and Hartman, G. L. 2015. From select agent to an established pathogen, the response to *Phakopsora pachyrhizi* in North America. Phytopathology 105:905-916.
Krupa, S., Bowersox, V., Claybrooke, R., Barnes, C.W., Szabo, L., Harlin, K. and Kurle, J. 2006. Introduction of Asian soybean rust urediniospores into the Midwestern United States—a case study. Plant Dis. 90:1254-1259.
Lambert, D. 1997. Zero-inflated Poisson regression, with an application to defects in manufacturing. Techometrics 34:1-14.
Livingstone, D. J., Manallack, D. T., and Tetko, I. V. 1997. Data modelling with neural networks: advantages and limitations. J. Comput. Aided Mol. Des. 11:135-142.
Maddox, R.L. 1870. On an apparatus for collecting atmospheric particles. Monthly Micros. J. **3**: 286–290.
Maddox, R.L. 1871. Observations on the use of the aeroconiscope, or air-dust collecting apparatus. Monthly Micros. J. **5**:45–49.
Matsuki, K., Kuperman, V., and Van Dyke, J. A. 2016. The Random forests statistical technique: An examination of its value for the study of reading. Sci. Stud. Read. 20:20-33.
Mehra, L. K., Cowger, C., Gross, K., and Ojiambo, P. S. 2016. Predicting pre-planting risk of *Stagonospora nodorum* blotch in winter wheat using machine learning models. Front. Plant Sci. 7 doi: 10.3389/fpls.2016.00390.
Melching, J. S., Bromfield, K. R., and Kingsolver, C. H. 1979. Infection, colonization, and uredospore production on Wayne soybean by four cultures of *Phakopsora pachyrhizi*, the cause of soybean rust. Phytopathology 69:1262-1265.
Miles, M. R., Levy, C., Morel, W., Mueller, T., Steinlage, T., van Rij, N., Frederick, R. D., and Hartman, G. L. 2007. International fungicide efficacy trials for the management of soybean rust. Plant Dis. 91:1450-1458.
Mundt, C. C., Wallace, L. D., Allen, T. W., Hollier, C. A., Kemerait, R. C., Sikora, E. J. 2013. Initial epidemic area is strongly associated with the yearly extent of soybean rust spread in North America. Biological Invasions 15:1431-1438.

NOAA. National Oceanic and Atmospheric Administration. Atmospheric Research Laboratory. NOAA ARL HYSPLIT Model. Online: http://www.arl.noaa.gov/ready/hysplit4.html. 2006.

Olatinwo, R. O., Paz, J. O., Brown, S. L., Kemerait, R. C., Culbreath, A. K., Beasley, J. P., and Hoogenboom, G. 2008. A predictive model for spotted wilt epidemics in peanut based on local weather conditions and the Tomato spotted wilt virus risk. Phytopathology 98:1066-1074.

Olden, J. D., Lawler, J. J., and Poff, N. L. 2008. Machine learning methods without tears: A primer for ecologists. Q. Rev. Biol. 83:171-193.

Oyeyemi, G. M., Ogunjobi, E. O. and Folorunsho, A. I., 2015. On performance of shrinkage methods–a Monte Carlo Study. Int. J. Statistics and App. 5:72-76

Pan, Z., Yang, X. B., Pivonia, S., Xue, L., Pasken, R., and Roads, J. 2006. Long-term prediction of soybean rust entry into the continental United States. Plant Dis. 90:840-846.

Paul, P. A., and Munkvold, G. P. 2005. Regression and artificial neural network modeling for the prediction of gray leaf spot of maize. Phytopathology 95:388-396.

R Development Core Team. 2008. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http: //www.R-project.org/.

Rapilly, F. 1979. Yellow rust epidemiology. Annual Review of Phytopathology 17:59-63.

Recknagel, F. 2001. Applications of machine learning to ecological modelling. Ecol. Model 146:303–310.

Rosso, P. H., and Hansen, E. M. 2003. Predicting Swiss needle cast disease distribution and severity in young Douglas-fir plantations in coastal Oregon. Phytopathology 93:790-798.

Savary, S., and Janeau, J. L.1986. Rain-induced dispersal of *Puccinia arachidis* studied by means of a rainfall simulator. Neth. J. Plant Pathol. 92:163-174.

Sikora, E. J., Allen, T. W., Wise, K. A., Bergstrom, G., Bradley, C. A., Bond, J., Brown-Rytlewski, D., Chilvers, M., Damicone, J., DeWolf, E., Dorrance, A., Dufault, N., Esker, P., Faske, T. R., Giesler, L., Goldberg, N., Golod, J., Gomez, I. R. G., Grau, C., Grybauskas, A., Franc, G., Hammerschmidt, R., Hartman, G. L., Henn, A., Hershman, D., Hollier, C., Isakeit, T., Isard, S., Jacobson, B., Jardine, D., Kemerait, B., Koenning, S., Langham, M., Malvick, D., Markell, S., Marois, J. J., Monfort, S., Mueller, D., Mueller, J., Mulroony, R., Newman, M., Osborne, L., Padgett, G. B., Ruden, B. E., Rupe, J., Schneider, R., Schwartz, H., Shaner, G., Singh, S., Stromberg, E., Sweets, L., Tenuta, A., Vaiciunas, S., Yang, X. B., Young-Kelly, H., and Zidek, J. 2014. A coordinated effort to manage soybean rust in North America: a success story in soybean disease monitoring. Plant Dis. 98:864-875.

Schneider, R. W., Hollier, C. A., Whitman, H. K., Palm, M. E., McKemy, J. M., Hernandez, J. R., Levy, L., and Devries-Paterson, R. 2005. First report of soybean rust caused by *Phakopsora pachyrhizi* in the continental United States. Plant Dis. 89:774.

Siroky D.S. 2009. Navigating random forests and related advances in algorithmic modeling. Statist. Surv. 3:147–163.

Tan, Y., Shan, Z., Shen, M., Yu, Z., Chang, R., Sun, J., Luo, Y., and Xiao, S. Evaluation of soybean germplasm of china for resistance to soybean rust. Soybean Sci. 16:205-209.

Tschanz, A. T., and Wang, T. C. 1980. Soybean rust development and apparent infection rates at five locations in Taiwan. Prot. Ecol. 2:247-250.

Uddin, W., Serlemitsos, K., and Viji, G. 2003. A temperature and leaf wetness duration-based model for prediction of gray leaf spot of perennial ryegrass turf. Phytopathology 93:336-343.

USDA. 2017. Pest Information Platform for Extension and Education. (Online) Available at http://www.sbrusa.net/. Accessed on Feb. 1, 2017.

Yang, X.B. 2006. Framework development in plant disease risk assessment and its application. European Journal of Plant Pathology 115:25-34.

Yorinori, J. T., Paiva, W. M., Frederick, R. D., Costamilan, L. M., Bertagnolli, P. F., Hartman, G. L., Godoy, C. V., and Nunes, J., Jr. 2005. Epidemics of soybean rust (*Phakopsora pachyrhizi*) in Brazil and Paraguay from 2001-2003. Plant Dis. 89:675-677.

APPENDIX A. SUPPLEMENTSRY TABLES AND FIGURES FOR PREDICTION OF SHORT-DISTANCE AERIAL MOVEMENT OF *PHAKOPSORA PACHYRHIZI* UREDINIOSPORES USING MACHINE LEARNING

TABLES

Table A.1. Analysis of variance (ANOVA) for the means of predicting variables under 'always-zero' group and 'non-zero' group in the passive trap dataset. variables with significantly different means were included in the inflation component of the zero-inflated Poisson model.

Predicting variables	Zero counts	Non-zero counts	P value[z]
Mean Daily Temperature	73.17	68.18	0.037*
Mean Daily Relative Humidity	83.85	81.44	0.56
Mean Daily Wind Speed	1.39	1.74	0.042*
Mean Wind Direction	215.59	220.58	0.76
Distance from Source	41.95	32.55	0.084
Minimum Wind Speed	3.92	4.49	0.06
Maximum Wind Speed	5.3	5.8	0.011*
Degree between trap and wind speed	88.05	91.78	0.49
Trap Position	149.25	158.02	0.67
Mean Daily Rainfall	0.13	0.09	0.19

[z] *refers to significant difference at $p < 0.05$

Table A.2. The Poisson count model component and zero-inflation component in the zero-inflation Poisson model. The table is showing the coefficients for each of the count predicting variables and the logit coefficients for the variable predicting excess zero along with standard errors, z-scores, and *p*-values for the coefficients.

Count model coefficients (Poisson with log link):				
	Estimate	Std. Error	z value	Pr>|t|[z]
(Intercept)	15.61981868	0.016815355	928.90212	<0.0001***
Mean Daily Temperature	-0.046443393	0.000269455	-172.3604	<0.0001***
Mean Daily Relative Humidity	-0.019355349	0.000150939	-128.23254	<0.0001***
Mean Daily Wind Speed	0.504454881	0.002485632	202.94836	<0.0001***
Mean Daily Wind Direction	-0.010359992	2.59866E-05	-398.66693	<0.0001***
Minimum Wind Speed	-0.002963668	0.000811121	-3.65379	0.00026***
Distance	-0.022919075	5.75963E-05	-397.92596	<0.0001***
Zero-inflation model coefficients (binomial with logit link):				
	Estimate	Std. Error	z value	Pr>|t|
(Intercept)	0.22729205	0.30309002	0.74992	0.453305
Max Daily Wind Speed	-0.11015005	0.05146499	-2.14029	0.032331 *

[z] *refers to significant difference at p <0.0001

FIGURES

Fig. A.1. Example of passive (left) and active (right) traps set up to capture *Phakopsora pachyrhizi* urediniospore from infected plants in the field.

Fig. A.2. Examples of a circular histograms depicting spore counts at Attapulgus on 6 October 2005. Each bar represents the spores per slide at one spore trap. A bar's angle represents a trap's position relative to the infected plot (0˚=north), and its magnitude represents spores/slide at (i) the 15 m distance and (ii) the 61 m distance. Wind speed and direction are depicted in (iii) the arrow plot. Each arrow represents mean wind speed and direction for 1 hr during active spore production and release (0700-1900 h). Each arrow's angle shows the direction toward which the wind was moving during an hour, and its magnitude represents the wind speed (km/h).

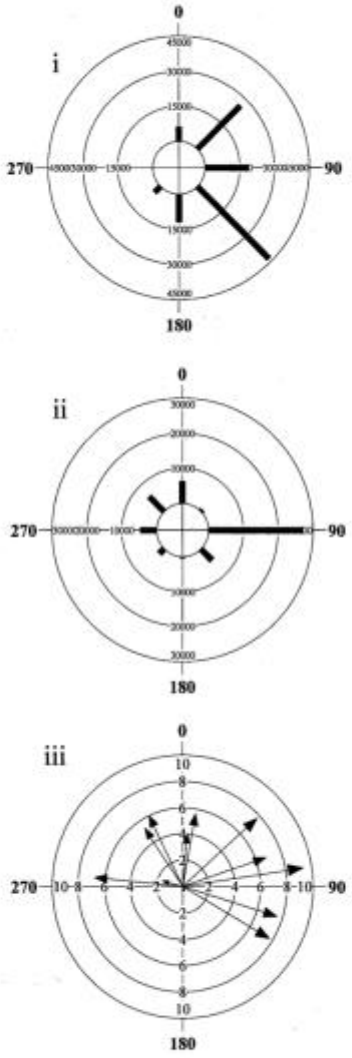


Fig. A.3. Performance profile to determine the optimal LASSO penalty tuning parameter lambda based on 10 repeats of 5-fold cross validation. The optimal penalty value corresponding to the best model is that which generated the smallest average MSE. A, Cross validation procedure on the passive trap dataset. B, Cross validation procedure on the active trap dataset.

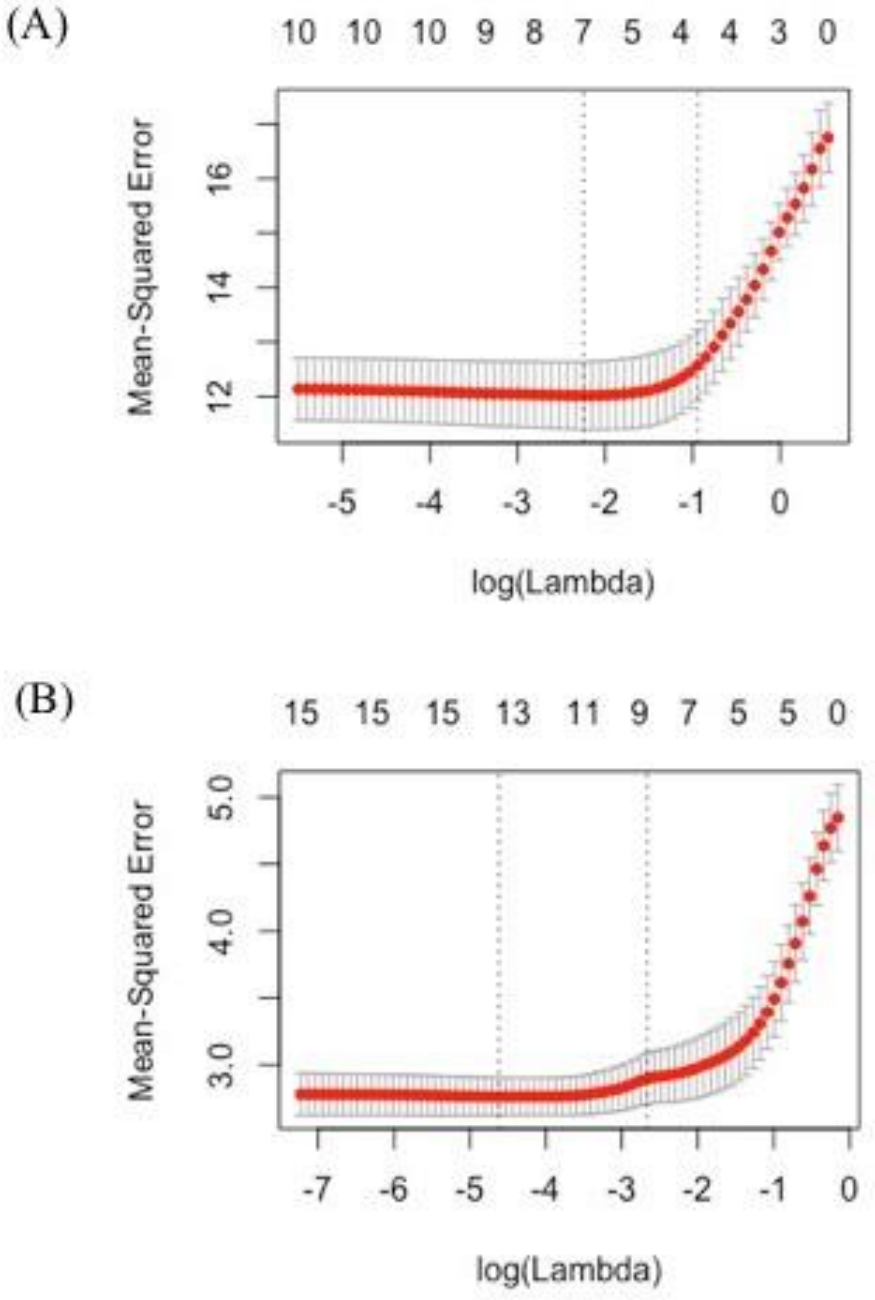


Fig. A.4. Performance profile to determine the optimal number of variables randomly sampled as candidates at each split for the random forest model based on 10 repeats of 5-fold cross validation. The optimal number corresponding to the best model is that which generated the smallest average MSE. A, Cross validation procedure on the passive trap dataset. B, Cross validation procedure on the active trap dataset.

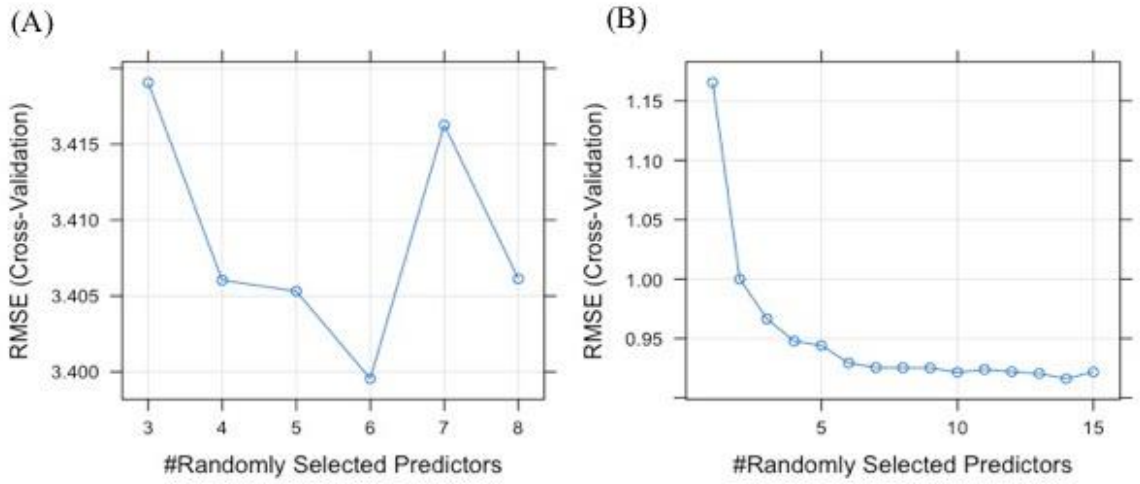
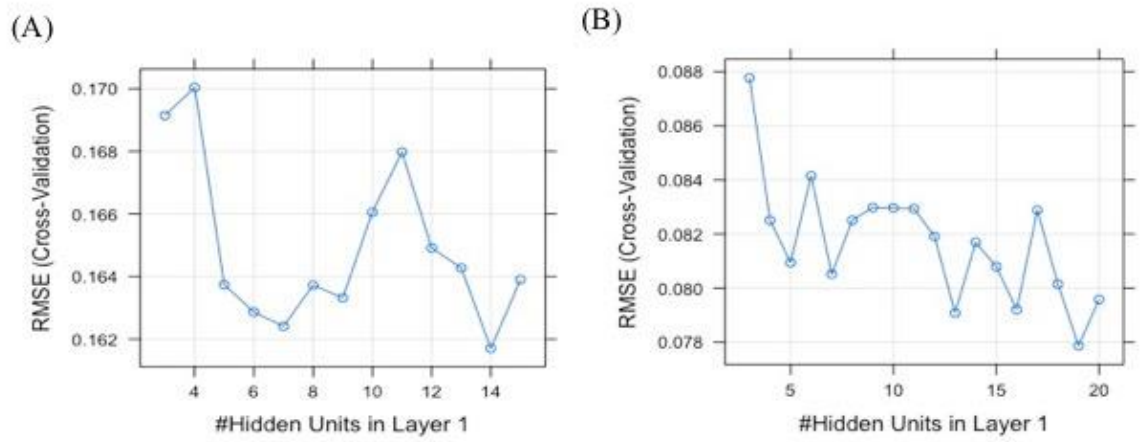


Fig. A.5. Performance profile to determine the optimal number of neurons in the first hidden layer based on 10 repeats of 5-fold cross validation for the neural network model. The optimal number of neurons corresponding to the best model is that which generated the smallest average MSE. A, Cross validation procedure on the passive trap dataset. B, Cross validation procedure on the active trap dataset.



APPENDIX B. GENOME-WIDE ASSOCIATION MAPPING FOR RESISTANCES TO SOYBEAN RUST AND SCLEROTINIA STEM ROT IN COMMON BEAN

ABSTRACT

Common bean (*Phaseolus vulgaris* L.) is a close relative of soybean (*Glycine max* (L.) Merr.), and some soybean pathogens can also infect *P. vulgaris* such as soybean rust (SBR), and Sclerotinia stem rot (SSR). Genetic resistance is the most effective strategy for disease management for both *G. max* and *P. vulgaris*, and new resistance sources for these diseases in *P. vulgaris* are needed for development of resistant common bean varieties. Genome-wide association study (GWAS) is a powerful tool to detect genetic markers that are significantly associated with resistant loci, which may harbor candidate resistance genes. With recent advance in acquiring single nucleotide polymorphism (SNP) markers, GWAS is regarded to have higher genomic resolution and mapping efficiency comparing to bi-parental linkage mapping. The objective of this study is to provide a comprehensive understanding of disease resistance to SBR, SSR in the core collection of *P. vulgaris* kept by the United States Department of Agriculture of Agricultural Research Service. Genotyping-by-sequencing was used to acquire SNPs for 363 *P. vulgaris* accessions. A total of 84,416 SNPs were identified by genotyping-by-sequencing with minor allele frequencies above 0.05. No significant SNPs were identified for SBR and SRR.

INTRODUCTION

Common bean, (*Phaseolus vulgaris* L.) is one of the most important grain legume in the human diet and a major protein source in some developing countries (Broughton et al., 2003) . It represents 50% of the grain legumes consumed worldwide (McClean et al., 2004). Common

bean has two geographically isolated and genetically different gene pools, with the Mesoamerica gene pool domesticated in Mexico and the Andean gene pool domesticated in Central and South America (McClean et al., 2004; Singh et al., 1991).

Common bean and soybean (*Glycine max* (L.) Merr.) are two closely related members in the *Phaseoleae*, a clade within the economically important Papilionoideae legumes. Many of the important soybean diseases could also be threats to common bean. Sclerotinia stem rot (SSR) caused by the fungus *Sclerotinia sclerotiorum*, and soybean rust (SBR) caused by the fungus *Phakopsora pachyrhizi*, are two of the most destructive pathogens on soybean, and they were considered to be two important causes of soybean yield losses. Under favorable conditions, SSR and SBR can cause up to 100% and 75% yield loss to susceptible soybean (Hartman et al. 1998; Hoffman et al. 1998; Yorinori et al. 2005). SSR is one of the most devastating fungal diseases on common bean and seed yield and quality loss due to this pathogen can be as high as 100 % under favorable conditions (Singh and Schwartz, 2010). SBR has been reported in common bean cultivars and elite lines under controlled and natural field conditions in South Africa, the United States, Argentina, and Brazil and is likely to become a new threat to common bean (Du Preez et al. 2005; Lynch et al. 2006; Pastor-Corrales et al. 2007; Ivancovich et al. 2005; Souza et al. 2014).

Developing resistant varieties is the most practical way to manage the two diseases, and characterization and identification of quantitative trait loci (QTLs) has been a great asset that facilitates the development of novel varieties with resistance to diseases in many crops. The genetic basis of resistance to the two diseases has been studied intensively in soybean. Ten loci (*Rpp1-Rpp6, Rpp1-b, Rpp[PI 567068A], Rpp?* (Hyuuga), *and Rpp6907*) conferring resistance to SBR have been identified so far and mapped on four chromosomes through bi-parental mapping

in soybean (Chakraborty et al. 2009; Chen et al. 2015; Garcia et al. 2008; Harris et al. 2015; Hossain et al., 2014; Hyten et al., 2007; Hyten et al. 2009; King et al. 2015; McLean and Byth 1980; Monteros et al. 2007; Silva et al. 2008). Similarly, QTLs associated with partial resistance to SSR have also been identified in a large number of plant introductions (PIs) using a range of inoculation methods in soybean. At least 35 QTLs have been identified thus far in soybean that have been associated with resistance to SSR (Chung et al., 2008; Miklas et al.,2013; Mkwaila et al., 2011; Pérez-Vega et al., 2012; Soule et al., 2011).

In comparison, the genetic basis of SBR resistance in common bean is not known, and there is no mapping study for SBR resistance in common bean. In order to develop resistant common bean varieties to SBR and improve marker-based selection, it is important to identify markers associated with SBR resistance in common bean. As for SSR resistance, some QTL mapping studies have been done and QTLs associated with partial resistance to SSR were identified in a large number of PIs using a range of inoculation methods in common bean. However, due to the quantitative nature and low heritability of the physiological resistance and architectural avoidance (upright plant architecture with open canopy) (Miklas et al., 2004; Park et al., 2001), there is still need to map additional or different QTLs in common bean.

In common bean, efforts have been made to use GWAS to detect and characterize QTL conveying a number of traits. Shi et al. (2011) studied common bacterial blight resistance with 132 SNPs in 469 dry bean breeding lines; they identified and confirmed 12 significant SNPs that were co-localized with or close to the common bacterial blight QTLs identified previously in bi-parental linkage mapping, and they also reported two novel QTLs for common bacterial blight resistance. In another study, the genetic architecture of five agronomic traits including pod fiber, seeds per pod, plant type, growth habit, and days to flowering were investigated by GWAS with

233 AFLP, 105 SNP and 80 SSR markers in 66 common bean genotypes (Nemli et al. 2014). They reported 13 pod fiber associated QTLs, 11 plant type associated QTLs, 14 growth habit associated QTLs, 2 QTLs for seed per pod, and 22 QTLs associated to days to flowering, and among the 62 significant markers identified, five of them had mutual associations for four traits.

There is no association mapping for SBR resistance in common bean, and no common bean variety with full resistance to SSR has been developed due to the quantitative nature and complexity of SSR resistance. GWAS has been employed as an effective strategy to identify complex traits in multiple genetic backgrounds, therefore, GWAS was applied to study the genetic architecture of SBR, and SSR resistance in the core collection of common bean.

MATERIALS AND METHODS

Plant material and phenotyping for soybean rust (SBR) and Sclerotinia stem rot (SSR)

The Central/South American (CA/SA, 171 accessions representing accessions from the Mesoamerican and Andean gene pool), and Mexico (MA, 191 accessions representing accessions from the Mesoamerican gene pool) core collections of common bean with 363 PIs maintained at the USDA/ARS Western Regional Plant Introduction Station, Pullman, WA were obtained (Brick et al., 2006; McClean et al., 2012). The reference genome of common bean G19833 was obtained from the International Center for Tropical Agriculture and used in this study (Schmutz et al., 2014).

For evaluating the resistance of the 363 common bean accessions to *S. sclerotiorum*, all the accessions were randomly planted in the 16 4X6-celled flats in soilless mix LC1 (Sun Gro Horticulture Inc., WA), with two seeds of each accession in each cell. The plants were maintained in the greenhouse at 25 ± 1ºC and 16-h photoperiod under 300 μmol m^{-2} s^{-1} light intensity. Seven-day-old seedlings were thinned to one plant per cell. All plants were inoculated

with the cut stem method described in a previously published paper three weeks after planting (Vuong et al., 2004). Plants were cut with a sterile razor blade 2 cm above the second node. The 200 µl pipette tips (Fisher Scientific) were pushed into the margin of the actively growing *S. sclerotiorum* cultures growing on potato dextrose agar. The pipette tip containing the agar disk with *S. sclerotiorum* mycelium was immediately placed over the cut stem and pushed down as far as possible to ensure greatest contact of the stem with the pathogen and to secure the tip on the stem. Inoculated plants were incubated in a greenhouse mist chamber with about 80% relative humidity overnight, and then plants were moved back to the greenhouse at 25 ± 1 ^{o}C with the same photoperiod and light intensity as before inoculation. Lesion length (cm) on each plant was measured a week after inoculation. The experiment was repeated once.

For evaluating the resistance of the 363 common bean accessions to *P. pachyrhizi*, all the accessions were also planted and arranged on the 16 4X6-celled flats, with two seeds of each accession in each cell. The plants were maintained in the greenhouse at 25 ± 1^{o}C and 16-h photoperiod under 300 µmol m^{-2} s^{-1} light intensity. Seven-day-old seedlings were thinned to one plant per cell, and all the seedlings were cut with only the uniforliate left. *P. pachyrhizi* isolate collected in Florida (FL07-1) in 2007 was multiplied as inoculum on susceptible soybean William 82, adult leaflets with *P. pachyrhizi* were harvested and placed in 0.01% Tween 20 (Vittal et al., 2014). The concentration was adjusted to 2.5×10^4 spores/ml. Three weeks after planting, all the accessions were inoculated with *P. pachyrhizi* by spraying the *P. pachyrhizi* urediniospores solution until run off. Inoculated plants were incubated in a greenhouse mist chamber with about 80% relative humidity overnight, and then plants were moved back to the greenhouse at 25 ± 1^{o}C with the same photoperiod and light intensity as before inoculation. Unifoliate leaves were collected for evaluation three week after inoculation.

A 1 cm^2 area was chosen on each leave on the area with most dense rust lesions. Lesion numbers was counting under a compound microscope with 20X magnification. The experiment was repeated once.

Genotyping of common bean collections

Genomic DNA was extracted from the emerging trifoliate leaf for each accession. Total genomic DNA was isolated using a standard cetyltrimethylammonium bromide protocol (Doyle and Doyle, 1987). DNA was quantified in 96-well plates using PicoGreen (Invitrogen, Carlsbad, CA) and DNA concentrations were normalized to 20 ng/ul. Restriction/ligation reactions were performed in 96-well plates using 500 ng of DNA from each individual, digestion with HindIII and BfaI (New England Biolabs, Ipswich, MA), and 0.1 μM and 10 μM of A1 and A2 adapters per well, respectively. The genomic libraries were pooled, column-cleaned using a PCR purification kit (Qiagen, Valencia, CA), and amplified for 12 cycles using Phusion DNA polymerase (IDT DNA). Average fragment size was estimated on a Bioanalyzer 2100 (Agilent, Santa Clara, CA) using a DNA1000 chip following a second column-cleaning. Pooled libraries were adjusted to 10 nmol and sequenced with 100-bp, single-end reads on an HiSeq2000 (Illumina, San Diego, CA). SNP calling was performed using Tassel 5 GBS v2 variant calling Pipeline IGST-GBS with G19833 as the reference genome (Glaubitz et al. 2014). A total of 212,281 "raw" SNPs and InDels were identified. Missing SNPs were imputed using BEAGLE version 4.1 (Browning and Browning, 2013). SNPs were further filtered with a minor allele frequency (MAF) and heterozygosity of 0.05.

Genome-wide association study (GWAS)

A total of 84,416 SNP markers were finally used for GWAS after imputation missing data and excluding ones with minor allele frequencies less than 5% or heterozygosity higher than

0.05. GWAS was performed using the Genomic Association and Prediction Integrated Tool (GAPIT) R package (Lipka et al., 2012). A kinship matrix (K) was calculated to determine relatedness among individuals (Zhang et al., 2010). The unified mixed linear model (MLM) which keeps kinship and population structure (K + P model) was used in this study to control population structure. SBR lesion numbers on the common bean PIs were log transformed to become normal. Significance was determined at false discovery rate (FDR) < 0.05.

RESULTS

Phenotypic analysis of resistance to SBR and SSR

P. pachyrhizi produced different types of lesions on the common bean collection (Fig. B.1), mainly differing in size. However, comparing to the tan lesion on the susceptible William 82, most of the lesions on common bean accessions are red-brown lesion with no or sparsely sporulating uredinia (Fig. B.2). There was a significant difference in SBR lesion numbers on common bean PIs (Table B.1) and the frequency distribution of soybean rust severity scores showed a continuous variation with population mean shifted towards susceptibility. A significant difference in SSR lesion length among common bean genotypes was also detected (Table B.2). The mean length of lesions covered a broad range, from as little as 0 mm to as long as 83 mm. susceptible checks had lesions averaging more than 63 mm.

SNP discovery and distribution

The identified 84,416 SNP markers were distributed over all 11 chromosomes with an average of 7674 SNP markers per chromosome (Table B.3). Chromosome 8 had the greatest number of SNPs (9333), and the chromosomes 9 had the lowest number of SNPs (5073).

Population structure

The population structure of the 363 common bean accessions was explored by principal com ponent analysis (PCA) using the 84,416 SNP markers and no distinct subpopulations among the 363 common bean PIs for any of the two diseases (Fig. B.3A), which is consistent with the result of the Baysian information criterion (BIC)-based model selection procedure. Therefore, no PC was included in the mixed linear model. A heatmap of the Kinship matrix with genetic relatednes s among the 363 common bean accessions indicated the two known gene pools (Andean gene po ol and Mesoamerican gene pool, Fig. B.3B) and three subpopulations in our bean collection. The refore, the Kinship was included in the mixed linear model for GWAS.

Linkage disequilibrium (LD) decay and candidate gene enrichment analyses.

The correlation coefficients (r^2) between all possible pairs of SNPs on each chromosome was calculated to establish LD relationships among loci. As expected, the r^2 value declined as the physical distance between the loci increased. LD decay for each chromosome was different (Table B.3). In addition, LD decay varied among all chromosomes, ranging from approximately 50 kb to 70 kb cut off $r^2 = 0.2$. The average LD decay for all chromosomes was estimated at approximately 60 kb.

GWAS for resistance to SBR and SSR

SSR lesion lengths and Log transformed SBR lesion numbers were normally distributed (Fig. B.4). A QQ plot of the GWAS for SSR and SBR resistance showed good false positive control (Fig. B.5). None of the tested SNPs were significantly associated with SBR and SSR resistance at 5% or 10% FDR (Fig. B.6).

DISCUSSION

Common bean shares multiple diseases with soybean, SBR and SSR are two top soybean diseases with highest damage potential on common bean. Resistances to the two diseases are very complex traits involving multiple genes. The identification of loci governing resistance not only help the genetic improvement of cultivars but also facilitates the identification of genes and the understanding of molecular mechanisms involved in the resistance process.

Although GWAS was proven a powerful tool for identifying and dissecting significant genomic regions harboring candidate genes with high genomic resolution and mapping efficiency, significant association between markers and traits is not guaranteed to be detected by GWAS. In our study, no significant SNPs were detected for resistance to SBR and SSR. The power of GWAS to identify a true association between a SNP and disease resistance is dependent on the effect size of the resistance allele and its allele frequency in the population (Korte and Farlow, 2013). Resistance to SSR in common bean is quantitatively inherited with low to moderate heritability, and the resistance mechanism is complex including physiological resistance and architectural avoidance. Therefore, the small allele effect size especially when confounding with genetic background may cause problem for the detection of significant association by GWAS. The negative GWAS result of soybean rust resistance suggested the complex genetic background of this trait. Traditional biparental mapping usually has advantage over GWAS in this situation by disentangling the population structure confounding effect and elevating the resistance allele frequency to intermediate level.

TABLES

Table B.1. Mean soybean rust lesion numbers on the PIs of the core collection of common bean. LSMeans difference of Tukey HSD test at α=0.05.

	Raw measurements			
PI	Rep1	Rep2	LSMeans[y]	Groups[z]
PI316016	27.3	28	27.65	a
PI313842	22	21	21.5	ab
PI511767	27.5	15	21.25	abc
PI209486	23.7	18	20.85	abcd
PI415886	20.5	20	20.25	abcde
PI181996	13.8	7	10.4	abcdef
PI189408	16.3	6	11.15	abcdef
PI196463	18.5	4.5	11.5	abcdef
PI198026	16.5	2	9.25	abcdef
PI198037	17.5	0	8.75	abcdef
PI207136	14.8	10	12.4	abcdef
PI208774	17	0	8.5	abcdef
PI269210	13.8	7	10.4	abcdef
PI282016	16.8	8	12.4	abcdef
PI290995	21.8	15	18.4	abcdef
PI293355	14.3	18	16.15	abcdef
PI297295	12.3	9	10.65	abcdef
PI299019	17.5	15	16.25	abcdef
PI307810	17.5	12	14.75	abcdef
PI309844	17.8	12	14.9	abcdef
PI313272	14.5	9	11.75	abcdef
PI313334	10.3	11	10.65	abcdef
PI313459	25	0	12.5	abcdef
PI313639	15.5	9.5	12.5	abcdef
PI313658	12.5	5	8.75	abcdef
PI313667	10.3	12	11.15	abcdef
PI313671	6.5	10.5	8.5	abcdef
PI313693	13.8	10.5	12.15	abcdef
PI313701	20.5	17.5	19	abcdef
PI316031	18.5	6	12.25	abcdef
PI317350	19	18	18.5	abcdef
PI319595	16.3	12.5	14.4	abcdef

[y] LSMeans of soybean rust lesion numbers on different common bean PIs calculated by Tukey HSD test
[z] LSMeans not connected with same letters are significantly different.

Table B1. (Cont.)

	Raw measurements			
PI	Rep 1	Rep 2	LSMeans[y]	Groups[z]
PI325618	18.8	15	16.9	abcdef
PI415936	11.3	6.5	8.9	abcdef
PI415950	13.5	9.5	11.5	abcdef
PI417641	21	15	18	abcdef
PI417653	17.8	0	8.9	abcdef
PI430210	15.5	9	12.25	abcdef
PI510574	9	14	11.5	abcdef
PI533312	12.5	6.5	9.5	abcdef
PI533332	12.8	8.5	10.65	abcdef
PI533510	21	0.5	10.75	abcdef
PI557483	13.5	9	11.25	abcdef
PI151407	9.3	0	4.65	bcdef
PI152208	6.5	1	3.75	bcdef
PI152311	8.5	1.5	5	bcdef
PI165422	5.3	0	2.65	bcdef
PI165462	8.3	1.5	4.9	bcdef
PI165466	3.8	2	2.9	bcdef
PI182004	7.3	0.5	3.9	bcdef
PI190078	10	0.5	5.25	bcdef
PI194574	11	0	5.5	bcdef
PI195402	12	2	7	bcdef
PI195801	7.7	0	3.85	bcdef
PI200967	5.3	0.5	2.9	bcdef
PI201004	4	3	3.5	bcdef
PI201010	6	0.5	3.25	bcdef
PI201296	3.5	3.5	3.5	bcdef
PI201329	13.8	0	6.9	bcdef
PI201343	14.5	0.5	7.5	bcdef
PI201369	5.5	0	2.75	bcdef
PI201370	7	0.5	3.75	bcdef
PI201480	7.5	0	3.75	bcdef
PI203920	6.5	0	3.25	bcdef
PI203924	6.5	0	3.25	bcdef
PI206223	3.8	1.5	2.65	bcdef

Table B.1. (Cont.)

	Raw measurements			
PI	Rep 1	Rep 2	LSMeans[y]	Groups[z]
PI207127	4	3	3.5	bcdef
PI207148	7.5	2	4.75	bcdef
PI207165	7	4	5.5	bcdef
PI207193	3.3	2	2.65	bcdef
PI207203	14	0	7	bcdef
PI207207	10	0	5	bcdef
PI207279	5	0.5	2.75	bcdef
PI207373	4.3	2	3.15	bcdef
PI207420	8	0	4	bcdef
PI207428	8.5	1.5	5	bcdef
PI209482	9.3	0	4.65	bcdef
PI209498	3.8	2	2.9	bcdef
PI224728	10	0.5	5.25	bcdef
PI241794	12	1	6.5	bcdef
PI260418	6.3	2	4.15	bcdef
PI263596	11.5	0	5.75	bcdef
PI268110	9.5	0	4.75	bcdef
PI269209	10	2	6	bcdef
PI293353	8.8	0	4.4	bcdef
PI307788	0.3	7.5	3.9	bcdef
PI307790	5.5	5	5.25	bcdef
PI307808	3.5	2	2.75	bcdef
PI308894	13	3	8	bcdef
PI308908	3.8	1.5	2.65	bcdef
PI309698	8.3	3	5.65	bcdef
PI309701	5.8	2.5	4.15	bcdef
PI309830	4.5	0.5	2.5	bcdef
PI309845	5	3	4	bcdef
PI309877	5	0	2.5	bcdef
PI310511	7.3	1	4.15	bcdef
PI310515	6	1	3.5	bcdef
PI310556	7.3	2.5	4.9	bcdef
PI310663	8.5	2	5.25	bcdef
PI310726	4.8	0.5	2.65	bcdef
PI310739	5.5	0	2.75	bcdef

Table B.1. (Cont.)

	Raw measurements			
PI	Rep 1	Rep 2	LSMeans[y]	Groups[z]
PI310826	6	0	3	bcdef
PI310829	7.5	3	5.25	bcdef
PI310850	5	2	3.5	bcdef
PI310865	11	4	7.5	bcdef
PI310886	4.8	5.5	5.15	bcdef
PI310891	6	2.5	4.25	bcdef
PI311794	3.3	5.5	4.4	bcdef
PI311843	5	5	5	bcdef
PI311853	3.5	2.5	3	bcdef
PI311940	4.5	1	2.75	bcdef
PI311947	8	0.5	4.25	bcdef
PI312017	14.5	0.5	7.5	bcdef
PI312083	5.5	0	2.75	bcdef
PI312098	8.8	4.5	6.65	bcdef
PI313270	10.5	1	5.75	bcdef
PI313297	4.5	0.5	2.5	bcdef
PI313328	4.8	0	2.4	bcdef
PI313357	8.3	4.5	6.4	bcdef
PI313394	3.5	2	2.75	bcdef
PI313408	3	2	2.5	bcdef
PI313412	9	1.5	5.25	bcdef
PI313429	6.3	0.5	3.4	bcdef
PI313444	8	0.5	4.25	bcdef
PI313445	6.5	0	3.25	bcdef
PI313458	3.3	1.5	2.4	bcdef
PI313483	4.8	0	2.4	bcdef
PI313490	10	0.5	5.25	bcdef
PI313532	3.8	1	2.4	bcdef
PI313537	4	1	2.5	bcdef
PI313571	5.3	1.5	3.4	bcdef
PI313572	7.8	0	3.9	bcdef
PI313598	6	1	3.5	bcdef
PI313608	4.3	1	2.65	bcdef
PI313613	5.3	0.5	2.9	bcdef
PI313630	4.8	3.5	4.15	bcdef

Table B.1. (Cont.)

	Raw measurements			
PI	Rep 1	Rep 2	LSMeans[y]	Groups[z]
PI313633	3.3	6	4.65	bcdef
PI313634	7.5	3	5.25	bcdef
PI313664	10	2	6	bcdef
PI313665	10.3	1	5.65	bcdef
PI313709	3.5	7	5.25	bcdef
PI313720	8.8	1	4.9	bcdef
PI313727	6.5	1	3.75	bcdef
PI313782	9.5	5.5	7.5	bcdef
PI313809	5	0	2.5	bcdef
PI313830	4	1	2.5	bcdef
PI313837	5.8	0	2.9	bcdef
PI313839	8	0	4	bcdef
PI313850	6	4	5	bcdef
PI317027	9.8	4.5	7.15	bcdef
PI318694	0	9	4.5	bcdef
PI319554	6.5	2.5	4.5	bcdef
PI319592	3.3	2.5	2.9	bcdef
PI319618	7.5	0	3.75	bcdef
PI319619	16	0.5	8.25	bcdef
PI325614	4.7	0.5	2.6	bcdef
PI325626	6.3	0.5	3.4	bcdef
PI325630	7	2	4.5	bcdef
PI325635	6.5	1	3.75	bcdef
PI325642	5.5	0	2.75	bcdef
PI325684	5	0	2.5	bcdef
PI325685	9.5	0	4.75	bcdef
PI325687	5	0	2.5	bcdef
PI325721	9.3	0	4.65	bcdef
PI325722	7.5	0	3.75	bcdef
PI346955	7.5	0	3.75	bcdef
PI346960	11	0	5.5	bcdef
PI355419	10.8	1	5.9	bcdef
PI387862	15.5	1	8.25	bcdef
PI387865	4	1	2.5	bcdef
PI399169	8	0.5	4.25	bcdef

Table B.1. (Cont.)

	Raw measurements			
PI	Rep 1	Rep 2	LSMeans[y]	Groups[z]
PI415913	10.3	3	6.65	bcdef
PI415949	8.3	5	6.65	bcdef
PI415955	9.5	1	5.25	bcdef
PI415975	9	2	5.5	bcdef
PI415986	6	4.5	5.25	bcdef
PI415987	5.3	1	3.15	bcdef
PI417621	4	1	2.5	bcdef
PI417622	4	1	2.5	bcdef
PI417628	3.8	2	2.9	bcdef
PI417630	6.5	2	4.25	bcdef
PI417633	9	4	6.5	bcdef
PI417634	7	3	5	bcdef
PI417645	6	2	4	bcdef
PI417647	8.5	0.5	4.5	bcdef
PI417654	6	4	5	bcdef
PI417657	7.5	3	5.25	bcdef
PI417679	11	3	7	bcdef
PI417707	1	4.5	2.75	bcdef
PI417708	9	4.5	6.75	bcdef
PI417739	5	3.5	4.25	bcdef
PI417742	6	0.5	3.25	bcdef
PI417743	7.5	0	3.75	bcdef
PI430167	7	1	4	bcdef
PI430201	4	1	2.5	bcdef
PI430204	11.3	1	6.15	bcdef
PI449389	2.3	2.5	2.4	bcdef
PI449422	3	2	2.5	bcdef
PI451885	4.5	5.5	5	bcdef
PI451906	2.5	5	3.75	bcdef
PI451917	5	0.5	2.75	bcdef
PI451921	6.5	2	4.25	bcdef
PI476751	2.3	3	2.65	bcdef
PI512003	7	0	3.5	bcdef
PI533249	3.8	1	2.4	bcdef
PI533259	7.5	3	5.25	bcdef

Table B.1. (Cont.)

	Raw measurements			
PI	Rep 1	Rep 2	LSMeans[y]	Groups[z]
PI533363	9	5	7	bcdef
PI533373	2.8	2	2.4	bcdef
PI533428	7	0	3.5	bcdef
PI533475	11.3	4.5	7.9	bcdef
PI533476	7.8	1	4.4	bcdef
PI533484	8	0.5	4.25	bcdef
PI533502	6	0	3	bcdef
PI533584	8.5	1	4.75	bcdef
PI535395	3.8	2.5	3.15	bcdef
PI197031	4.3	0	2.15	cdef
PI207154	3.5	1	2.25	cdef
PI209479	4.5	0	2.25	cdef
PI309823	4.3	0	2.15	cdef
PI311999	4	0.5	2.25	cdef
PI313373	4	0.5	2.25	cdef
PI313486	3.5	1	2.25	cdef
PI313674	1	3.5	2.25	cdef
PI533420	3.8	0.5	2.15	cdef
PI165423	3	0.5	1.75	def
PI182000	4	0	2	def
PI189407	4	0	2	def
PI201360	3	1	2	def
PI203958	2.3	1.5	1.9	def
PI207253	2.5	1	1.75	def
PI290990	2.8	1	1.9	def
PI307816	2.8	1	1.9	def
PI307823	3.5	0	1.75	def
PI310883	1.3	2.5	1.9	def
PI310915	2	2	2	def
PI312031	4	0	2	def
PI312052	3	1	2	def
PI313237	2	2	2	def
PI313583	3	1	2	def
PI313597	2.5	1	1.75	def
PI313733	3.5	0	1.75	def

Table B.1. (Cont.)

	Raw measurements			
PI	Rep 1	Rep 2	LSMeans[y]	Groups[z]
PI319607	2.5	1	1.75	def
PI319636	0.3	3.5	1.9	def
PI319683	3	1	2	def
PI325691	4	0	2	def
PI415900	3	0.5	1.75	def
PI416713	2.5	1.5	2	def
PI417616	1	3	2	def
PI417667	1.8	2	1.9	def
PI417778	2	1.5	1.75	def
PI533286	2.5	1	1.75	def
PI533311	3.3	0.5	1.9	def
PI533313	3	0.5	1.75	def
PI533316	3.5	0.5	2	def
PI533498	2	1.5	1.75	def
PI150957	2.5	0	1.25	ef
PI200956	3	0	1.5	ef
PI201388	2.3	0	1.15	ef
PI202834	3	0	1.5	ef
PI202835	2.3	0	1.15	ef
PI203921	2	1	1.5	ef
PI203936	2.8	0.5	1.65	ef
PI207182	2	1	1.5	ef
PI207216	2.5	0.5	1.5	ef
PI207336	3	0	1.5	ef
PI207389	1.5	1	1.25	ef
PI207443	2	1	1.5	ef
PI304113	2	1	1.5	ef
PI308898	2.5	0	1.25	ef
PI309715	2.5	0	1.25	ef
PI309759	2.3	1	1.65	ef
PI309787	3	0	1.5	ef
PI309825	3	0	1.5	ef
PI310561	1.3	1.5	1.4	ef
PI310586	2.5	0	1.25	ef
PI310599	2.8	0	1.4	ef

Table B.1. (Cont.)

	Raw measurements			
PI	Rep 1	Rep 2	LSMeans[y]	Groups[z]
PI310690	3	0	1.5	ef
PI310814	2.5	0	1.25	ef
PI311962	2	0.5	1.25	ef
PI312064	1.8	1.5	1.65	ef
PI312090	1.8	0.5	1.15	ef
PI313366	3	0	1.5	ef
PI313440	2	0.5	1.25	ef
PI313470	2.5	0	1.25	ef
PI313495	2.3	0	1.15	ef
PI313501	2.3	0	1.15	ef
PI313531	2.8	0.5	1.65	ef
PI313609	2.3	1	1.65	ef
PI319573	1	1.5	1.25	ef
PI319684	2.5	0	1.25	ef
PI326106	2	0.5	1.25	ef
PI326110	2	1	1.5	ef
PI417716	0.5	2	1.25	ef
PI417721	1.3	1	1.15	ef
PI417731	0.5	2	1.25	ef
PI430200	2.8	0	1.4	ef
PI430206	2.5	0	1.25	ef
PI533281	2.8	0	1.4	ef
PI150409	0.8	0	0.4	f
PI165455	1.5	0	0.75	f
PI189016	0	0	0	f
PI201324	2	0	1	f
PI201354	1.5	0	0.75	f
PI201387	0.5	1	0.75	f
PI207180	1.5	0	0.75	f
PI207186	1	0.5	0.75	f
PI207300	1.5	0	0.75	f
PI209491	2	0	1	f
PI224715	1.3	0	0.65	f
PI224718	1.3	0.5	0.9	f
PI288016	0.5	1	0.75	f

Table B.1. (Cont.)

	Raw measurements			
PI	Rep 1	Rep 2	LSMeans[y]	Groups[z]
PI307791	2	0	1	f
PI307806	2	0	1	f
PI307820	1.5	0.5	1	f
PI309700	1.5	0	0.75	f
PI309810	1	0	0.5	f
PI309827	1	1	1	f
PI309837	1.8	0	0.9	f
PI309857	1.5	0.5	1	f
PI310546	0	0.5	0.25	f
PI310611	0.8	0	0.4	f
PI310660	1.5	0	0.75	f
PI310778	0	0.5	0.25	f
PI310828	0	0.5	0.25	f
PI310836	0.8	0.5	0.65	f
PI311942	1.8	0	0.9	f
PI311967	1	0	0.5	f
PI311974	0.8	0.5	0.65	f
PI312016	0	1	0.5	f
PI312018	1.5	0	0.75	f
PI313254	0	0.5	0.25	f
PI313386	0	0.5	0.25	f
PI313512	1.3	0	0.65	f
PI313592	1.5	0	0.75	f
PI313749	1.3	0	0.65	f
PI313833	2	0	1	f
PI313835	0	0.5	0.25	f
PI313847	1.8	0	0.9	f
PI319587	1.5	0	0.75	f
PI319640	1.3	0.5	0.9	f
PI319674	2	0	1	f
PI325653	0	1	0.5	f
PI325676	1	0	0.5	f
PI325732	1.5	0.5	1	f
PI325750	0.3	0	0.15	f
PI345576	1.3	0	0.65	f

Table B.1. (Cont.)

	Raw measurements			
PI	Rep 1	Rep 2	LSMeans[y]	Groups[z]
PI345581	0	0	0	f
PI406940	0	0.5	0.25	f
PI417754	1.5	0	0.75	f
PI417780	1	0.5	0.75	f
PI417784	1	0	0.5	f
PI417790	0	0.5	0.25	f
PI451889	1.5	0	0.75	f
PI533277	1.3	0.5	0.9	f
PI533299	2	0	1	f
PI533432	0	1	0.5	f
PI533437	0.5	0.5	0.5	f
PI533528	1	0	0.5	f
PI533545	0	2	1	f
PI533561	0.5	1	0.75	f

Table B.2. Mean Sclerotenia stem rot lesion length on the PIs of the core collection of common bean. LSMeans difference of Tukey HSD test at α=0.05.

	Raw measurements			
PI	Rep1	Rep2	LSMeans[y]	Groups[z]
PI309700	9	9.1	9.05	a
PI309825	8.7	8.7	8.7	ab
PI325684	8.5	8.5	8.5	abc
PI207186	8.5	8.4	8.45	abcd
PI319674	5.6	10.9	8.25	abcde
PI207279	8.7	7.1	7.9	abcdef
PI150409	2.5	3.5	3	abcdefg
PI150957	3.5	0.2	1.85	abcdefg
PI151407	4	6	5	abcdefg
PI152208	4.9	5.5	5.2	abcdefg
PI152311	4.5	5.6	5.05	abcdefg
PI165422	3.8	5.2	4.5	abcdefg
PI165423	3.5	3.6	3.55	abcdefg
PI165455	2.8	4.8	3.8	abcdefg
PI165462	0.5	2.6	1.55	abcdefg
PI181996	5.9	0	2.95	abcdefg
PI182000	8.5	6	7.25	abcdefg
PI182004	0	4.5	2.25	abcdefg
PI189016	4.5	4.6	4.55	abcdefg
PI189407	4	5.5	4.75	abcdefg
PI189408	0	3	1.5	abcdefg
PI190078	5.5	4	4.75	abcdefg
PI194574	4.5	3.7	4.1	abcdefg
PI195402	5.5	5.5	5.5	abcdefg
PI195801	6	1	3.5	abcdefg
PI196463	4.5	2.1	3.3	abcdefg
PI197031	1.5	3.2	2.35	abcdefg
PI198026	4.9	4.9	4.9	abcdefg
PI198037	2.5	5.5	4	abcdefg
PI200956	1.4	2.2	1.8	abcdefg
PI200967	6	3.8	4.9	abcdefg
PI201004	1.3	2.6	1.95	abcdefg
PI201010	2.8	2.3	2.55	abcdefg
PI201296	8.2	3.8	6	abcdefg

[y] LSMeans of soybean rust lesion numbers on different common bean PIs calculated by Tukey HSD test

[z] LSMeans not connected with same letters are significantly different.

Table B.2. (Cont.)

	Raw measurements			
PI	Rep 1	Rep 2	LSMeans[y]	Groups[z]
PI201324	7.6	7	7.3	abcdefg
PI201329	8.6	5.5	7.05	abcdefg
PI201343	4.1	2	3.05	abcdefg
PI201354	3.5	5.2	4.35	abcdefg
PI201360	7.5	5.8	6.65	abcdefg
PI201369	2.6	4.7	3.65	abcdefg
PI201370	6.1	6.5	6.3	abcdefg
PI201387	5	4.2	4.6	abcdefg
PI201388	6	6.2	6.1	abcdefg
PI201480	5.2	5.9	5.55	abcdefg
PI202834	2.1	5.5	3.8	abcdefg
PI202835	2.2	1.2	1.7	abcdefg
PI203920	5	2.6	3.8	abcdefg
PI203921	7.5	4.5	6	abcdefg
PI203924	4	5.4	4.7	abcdefg
PI203936	6	4.5	5.25	abcdefg
PI203958	8.5	6.5	7.5	abcdefg
PI206223	4.1	1.2	2.65	abcdefg
PI207136	0.2	3	1.6	abcdefg
PI207154	0.2	2.7	1.45	abcdefg
PI207165	0.2	3	1.6	abcdefg
PI207180	0.2	4.6	2.4	abcdefg
PI207193	7.2	5.8	6.5	abcdefg
PI207203	1	2.8	1.9	abcdefg
PI207207	3.8	4	3.9	abcdefg
PI207216	3.5	3.3	3.4	abcdefg
PI207300	2.8	3.4	3.1	abcdefg
PI207336	3.5	6.5	5	abcdefg
PI207373	4.7	4.2	4.45	abcdefg
PI207389	3.2	2.6	2.9	abcdefg
PI207420	4	5	4.5	abcdefg
PI207428	3.1	1.7	2.4	abcdefg
PI207443	6	5	5.5	abcdefg
PI208774	2	3.5	2.75	abcdefg
PI209479	6	3.5	4.75	abcdefg

Table B.2. (Cont.)

	Raw measurements			
PI	Rep 1	Rep 2	LSMeans[y]	Groups[z]
PI209482	7	2.9	4.95	abcdefg
PI209491	3.4	5.5	4.45	abcdefg
PI209498	0	3	1.5	abcdefg
PI224715	7	1.2	4.1	abcdefg
PI224718	5	4.3	4.65	abcdefg
PI224728	1	3.9	2.45	abcdefg
PI241794	5.3	5.6	5.45	abcdefg
PI260418	3	4.5	3.75	abcdefg
PI268110	3.2	4.5	3.85	abcdefg
PI269209	7.9	7	7.45	abcdefg
PI269210	6.5	6.7	6.6	abcdefg
PI282016	3.6	4.2	3.9	abcdefg
PI288016	1.2	4.1	2.65	abcdefg
PI290990	3.8	0.5	2.15	abcdefg
PI290995	1.9	6	3.95	abcdefg
PI293353	0.2	2.6	1.4	abcdefg
PI293355	2.5	6	4.25	abcdefg
PI297295	0.5	2.6	1.55	abcdefg
PI299019	0.5	4.5	2.5	abcdefg
PI304113	7	7.6	7.3	abcdefg
PI307788	5.5	1.7	3.6	abcdefg
PI307790	6.5	0	3.25	abcdefg
PI307806	2.7	0.7	1.7	abcdefg
PI307808	7.5	7	7.25	abcdefg
PI307810	4.6	2.5	3.55	abcdefg
PI307816	2.5	0.6	1.55	abcdefg
PI307820	3	8.5	5.75	abcdefg
PI307823	2.6	1.2	1.9	abcdefg
PI308894	4	4.7	4.35	abcdefg
PI308898	3.5	5.5	4.5	abcdefg
PI308908	3.3	4.4	3.85	abcdefg
PI309698	4.2	5.2	4.7	abcdefg
PI309715	1.5	5.2	3.35	abcdefg
PI309759	4.5	4.5	4.5	abcdefg
PI309787	2	2	2	abcdefg

Table B.2. (Cont.)

	Raw measurements			
PI	Rep 1	Rep 2	LSMeans[y]	Groups[z]
PI309810	0.8	2.9	1.85	abcdefg
PI309823	8.5	4.2	6.35	abcdefg
PI309827	5.5	3.1	4.3	abcdefg
PI309830	6.1	3.5	4.8	abcdefg
PI309837	3.2	5.5	4.35	abcdefg
PI309844	3.9	3	3.45	abcdefg
PI309857	6.2	8.3	7.25	abcdefg
PI309877	6	2.5	4.25	abcdefg
PI310515	1	4.1	2.55	abcdefg
PI310556	5.7	4.1	4.9	abcdefg
PI310561	5.5	4.2	4.85	abcdefg
PI310586	2.6	2.5	2.55	abcdefg
PI310599	5.2	5.2	5.2	abcdefg
PI310611	4.4	5.6	5	abcdefg
PI310660	3	5.6	4.3	abcdefg
PI310663	5.6	3	4.3	abcdefg
PI310690	2.1	3.5	2.8	abcdefg
PI310726	2	6.9	4.45	abcdefg
PI310778	4.1	3.4	3.75	abcdefg
PI310826	3	2.5	2.75	abcdefg
PI310828	2.9	2.1	2.5	abcdefg
PI310829	7	5.1	6.05	abcdefg
PI310836	3.5	2.2	2.85	abcdefg
PI310850	3.5	3	3.25	abcdefg
PI310865	0.2	3.6	1.9	abcdefg
PI310883	5.6	1.5	3.55	abcdefg
PI310886	3	5.7	4.35	abcdefg
PI310891	3.4	3.2	3.3	abcdefg
PI310915	5.4	4.5	4.95	abcdefg
PI311794	7	2.5	4.75	abcdefg
PI311940	8.5	4.5	6.5	abcdefg
PI311942	3.5	5.5	4.5	abcdefg
PI311947	3.8	6.9	5.35	abcdefg
PI311962	3.5	1	2.25	abcdefg
PI311967	2	5	3.5	abcdefg

Table B.2. (Cont.)

	Raw measurements			
PI	Rep 1	Rep 2	LSMeans[y]	Groups[z]
PI311974	2	3.7	2.85	abcdefg
PI311999	3.5	0.5	2	abcdefg
PI312016	0.2	3.5	1.85	abcdefg
PI312017	0	4	2	abcdefg
PI312018	0	4.7	2.35	abcdefg
PI312031	3	2.8	2.9	abcdefg
PI312052	4.5	3.4	3.95	abcdefg
PI312064	1	4.5	2.75	abcdefg
PI312083	4	0.5	2.25	abcdefg
PI312090	2.7	7.5	5.1	abcdefg
PI312098	3.7	2.5	3.1	abcdefg
PI313237	2.8	2.8	2.8	abcdefg
PI313254	4.3	2	3.15	abcdefg
PI313270	2.5	1.9	2.2	abcdefg
PI313272	2.2	2.9	2.55	abcdefg
PI313297	2.5	1.2	1.85	abcdefg
PI313328	6.2	6.5	6.35	abcdefg
PI313334	5.5	4.5	5	abcdefg
PI313357	2.6	5.6	4.1	abcdefg
PI313373	2	6.5	4.25	abcdefg
PI313386	7	0.4	3.7	abcdefg
PI313394	3.4	4	3.7	abcdefg
PI313408	2.5	2.2	2.35	abcdefg
PI313412	5	4.5	4.75	abcdefg
PI313429	0.5	3.5	2	abcdefg
PI313440	0.6	5	2.8	abcdefg
PI313444	3	0	1.5	abcdefg
PI313445	3	4.9	3.95	abcdefg
PI313458	2	3.5	2.75	abcdefg
PI313459	0.2	2.7	1.45	abcdefg
PI313470	6	5.8	5.9	abcdefg
PI313483	5.5	5.8	5.65	abcdefg
PI313486	7.5	7.2	7.35	abcdefg
PI313490	7	5.9	6.45	abcdefg
PI313495	8	7.2	7.6	abcdefg

Table B.2. (Cont.)

	Raw measurements			
PI	Rep 1	Rep 2	LSMeans[y]	Groups[z]
PI313501	3.5	3	3.25	abcdefg
PI313512	6.5	6.2	6.35	abcdefg
PI313531	6.3	2.5	4.4	abcdefg
PI313532	3	5.4	4.2	abcdefg
PI313537	2.9	2.9	2.9	abcdefg
PI313571	5	0	2.5	abcdefg
PI313572	5.5	4.9	5.2	abcdefg
PI313592	3.2	2.5	2.85	abcdefg
PI313597	7.5	3	5.25	abcdefg
PI313598	4.2	2.5	3.35	abcdefg
PI313608	6.5	6.1	6.3	abcdefg
PI313613	4.5	3	3.75	abcdefg
PI313630	3.5	2.1	2.8	abcdefg
PI313633	6	2.7	4.35	abcdefg
PI313634	4.7	5.9	5.3	abcdefg
PI313639	5.5	6.4	5.95	abcdefg
PI313658	2.2	3.2	2.7	abcdefg
PI313665	3.7	3	3.35	abcdefg
PI313671	3	5	4	abcdefg
PI313674	2.6	0.5	1.55	abcdefg
PI313693	2.4	0.5	1.45	abcdefg
PI313701	3.8	0.5	2.15	abcdefg
PI313709	6	6.1	6.05	abcdefg
PI313720	5.3	3.2	4.25	abcdefg
PI313727	7	6.8	6.9	abcdefg
PI313733	2.1	1.9	2	abcdefg
PI313749	7.2	6.7	6.95	abcdefg
PI313782	6	2	4	abcdefg
PI313809	7.5	5.1	6.3	abcdefg
PI313830	7	1.4	4.2	abcdefg
PI313835	1	5.6	3.3	abcdefg
PI313837	0.8	3	1.9	abcdefg
PI313839	6.8	4.9	5.85	abcdefg
PI313842	5.6	5.9	5.75	abcdefg
PI313850	2.9	3.2	3.05	abcdefg

Table B.2. (Cont.)

	Raw measurements			
PI	Rep 1	Rep 2	LSMeans[y]	Groups[z]
PI316031	4	3.6	3.8	abcdefg
PI317027	3.2	3.1	3.15	abcdefg
PI317350	2.6	2	2.3	abcdefg
PI318694	3.1	3	3.05	abcdefg
PI319554	2	3.5	2.75	abcdefg
PI319587	6.5	6.1	6.3	abcdefg
PI319592	6.5	5.3	5.9	abcdefg
PI319595	3	7.7	5.35	abcdefg
PI319607	5.1	4.9	5	abcdefg
PI319618	5.1	1.5	3.3	abcdefg
PI319619	5.2	5.9	5.55	abcdefg
PI319636	3.2	1.5	2.35	abcdefg
PI319683	3.5	1.7	2.6	abcdefg
PI319684	3.8	9.1	6.45	abcdefg
PI325614	6	7	6.5	abcdefg
PI325618	8	7	7.5	abcdefg
PI325626	6.7	5.6	6.15	abcdefg
PI325630	6.8	5.2	6	abcdefg
PI325635	8.2	3.6	5.9	abcdefg
PI325642	3.2	4.6	3.9	abcdefg
PI325653	1.5	3.5	2.5	abcdefg
PI325676	3.2	3	3.1	abcdefg
PI325685	4	1.5	2.75	abcdefg
PI325687	3	2.8	2.9	abcdefg
PI325691	2.5	1.8	2.15	abcdefg
PI325721	3.5	1.7	2.6	abcdefg
PI325732	6.6	3.1	4.85	abcdefg
PI325750	5.5	6	5.75	abcdefg
PI326106	5.2	2.6	3.9	abcdefg
PI326110	3	1.4	2.2	abcdefg
PI345576	4.1	2.4	3.25	abcdefg
PI345581	6.4	8.2	7.3	abcdefg
PI346955	3.2	3.2	3.2	abcdefg
PI346960	7	6.9	6.95	abcdefg
PI387862	11	3.6	7.3	abcdefg

Table B.2. (Cont.)

	Raw measurements			
PI	Rep 1	Rep 2	LSMeans[y]	Groups[z]
PI387865	3	0	1.5	abcdefg
PI399169	3	0	1.5	abcdefg
PI406940	7	5.7	6.35	abcdefg
PI415913	3.5	4.5	4	abcdefg
PI415936	5	0	2.5	abcdefg
PI415949	6.5	3.9	5.2	abcdefg
PI415987	0.2	4.2	2.2	abcdefg
PI417616	5.5	6	5.75	abcdefg
PI417621	3.2	4.6	3.9	abcdefg
PI417622	5	5	5	abcdefg
PI417628	5.6	4.5	5.05	abcdefg
PI417630	6.5	3.7	5.1	abcdefg
PI417633	4.8	5.2	5	abcdefg
PI417634	5.5	5.8	5.65	abcdefg
PI417641	4.5	3.9	4.2	abcdefg
PI417645	6.5	6	6.25	abcdefg
PI417647	2.5	1.5	2	abcdefg
PI417653	3.5	5	4.25	abcdefg
PI417657	3.7	6.5	5.1	abcdefg
PI417679	4.5	5	4.75	abcdefg
PI417708	6	6.3	6.15	abcdefg
PI417716	7.2	5.3	6.25	abcdefg
PI417721	3.1	4.9	4	abcdefg
PI417731	2.9	2.2	2.55	abcdefg
PI417739	6	4.6	5.3	abcdefg
PI417742	6	6.8	6.4	abcdefg
PI417743	7.5	6.5	7	abcdefg
PI417754	3.5	5.2	4.35	abcdefg
PI417778	4	3.5	3.75	abcdefg
PI417780	4	3.2	3.6	abcdefg
PI417784	7	6.8	6.9	abcdefg
PI417790	5.9	6	5.95	abcdefg
PI430200	6.5	3.7	5.1	abcdefg
PI430201	5	4.6	4.8	abcdefg
PI430204	6.5	7	6.75	abcdefg

Table B.2. (Cont.)

	Raw measurements			
PI	Rep 1	Rep 2	LSMeans[y]	Groups[z]
PI430206	3	4	3.5	abcdefg
PI430210	3	4.4	3.7	abcdefg
PI449389	3.4	4.4	3.9	abcdefg
PI449422	2.5	4.5	3.5	abcdefg
PI451885	6	3.5	4.75	abcdefg
PI451889	3	3	3	abcdefg
PI451906	1.2	2.2	1.7	abcdefg
PI451917	5.5	4	4.75	abcdefg
PI451921	4.7	6.6	5.65	abcdefg
PI476751	3.7	0.5	2.1	abcdefg
PI510574	2.7	6.5	4.6	abcdefg
PI511767	2.5	2.1	2.3	abcdefg
PI512003	2.7	2	2.35	abcdefg
PI533249	5.7	4	4.85	abcdefg
PI533259	3.5	2	2.75	abcdefg
PI533277	5	3.2	4.1	abcdefg
PI533281	3	2.5	2.75	abcdefg
PI533286	3.5	1	2.25	abcdefg
PI533299	3.6	3.1	3.35	abcdefg
PI533311	7.2	3.5	5.35	abcdefg
PI533312	5.8	7.4	6.6	abcdefg
PI533313	3.7	1.1	2.4	abcdefg
PI533316	3.5	3.2	3.35	abcdefg
PI533363	3.5	5.5	4.5	abcdefg
PI533373	7	4	5.5	abcdefg
PI533420	5.1	4.1	4.6	abcdefg
PI533428	4.6	8.5	6.55	abcdefg
PI533432	7.2	2.3	4.75	abcdefg
PI533437	1.5	2.9	2.2	abcdefg
PI533475	3.5	3.6	3.55	abcdefg
PI533476	6.2	3.1	4.65	abcdefg
PI533484	3.6	5	4.3	abcdefg
PI533498	1.1	2	1.55	abcdefg
PI533502	2.5	2.5	2.5	abcdefg
PI533510	3.2	4.2	3.7	abcdefg

Table B.2. (Cont.)

	Raw measurements			
PI	Rep 1	Rep 2	LSMeans[y]	Groups[z]
PI533528	5.6	4	4.8	abcdefg
PI533545	4.8	4	4.4	abcdefg
PI533561	2.1	3.2	2.65	abcdefg
PI533584	3.3	3.5	3.4	abcdefg
PI535395	2	2.1	2.05	abcdefg
PI557483	3.9	3.1	3.5	abcdefg
PI207182	0.8	1.6	1.2	bcdefg
PI310546	2	0.2	1.1	bcdefg
PI310814	0	2.5	1.25	bcdefg
PI313609	2.5	0	1.25	bcdefg
PI313833	1.5	1.1	1.3	bcdefg
PI415955	1	1.5	1.25	bcdefg
PI416713	0.5	1.5	1	bcdefg
PI417707	2.5	0	1.25	bcdefg
PI207127	0.5	1.3	0.9	cdefg
PI309845	0.2	1.7	0.95	cdefg
PI415900	0.2	1.5	0.85	cdefg
PI415975	0.3	1.5	0.9	cdefg
PI430167	1.5	0.2	0.85	cdefg
PI307791	0.5	1	0.75	defg
PI309701	0.5	1	0.75	defg
PI313667	1	0.5	0.75	defg
PI165466	0.2	1	0.6	efg
PI310739	0	1.2	0.6	efg
PI319640	0.9	0.3	0.6	efg
PI325722	0	1.1	0.55	efg
PI415950	0	1.2	0.6	efg
PI417667	1	0.2	0.6	efg
PI533332	1	0.2	0.6	efg
PI207148	0.2	0.2	0.2	fg
PI209486	0.5	0.5	0.5	fg
PI263596	0.5	0.5	0.5	fg
PI310511	0.2	0.8	0.5	fg
PI311843	0	0.8	0.4	fg
PI313366	0.6	0.2	0.4	fg

Table B.2. (Cont.)

	Raw measurements			
PI	Rep 1	Rep 2	LSMeans[y]	Groups[z]
PI313583	0.5	0.5	0.5	fg
PI313664	0	0.6	0.3	fg
PI313847	0.5	0.5	0.5	fg
PI316016	0.7	0	0.35	fg
PI319573	0.2	0.5	0.35	fg
PI415886	0.3	0.2	0.25	fg
PI415986	0.2	0.5	0.35	fg
PI417654	0	0.8	0.4	fg
PI207253	0	0.2	0.1	g
PI318703	0	0.2	0.1	g
PI355419	0	0	0	g

Table B.3. Linkage disequilibrium (LD) decay estimated for different soybean chromosomes

Chr No.[w]	Chr size (kb)	No. of SNPs[x]	LD decay (kb)[y]	SNPs needed[z]
1	52183.5	8571	70	745
2	49033.7	8559	60	817
3	52218.6	8549	60	870
4	45793.2	8247	65	705
5	40237.5	7313	65	619
6	31973.2	8600	65	492
7	51698.4	6289	60	826
8	59634.6	9333	50	1193
9	37399.6	5073	60	623
10	43213.2	7662	60	720
11	50203.6	9220	50	1004

[w]Chr No. – Chromosome number
[x]Number of SNPs used in present study.
[y]LD decay at $r^2 = 0.2$.
[z]Average SNPs needed was calculated by deviding chromosome size by LD decay.

FIGURES

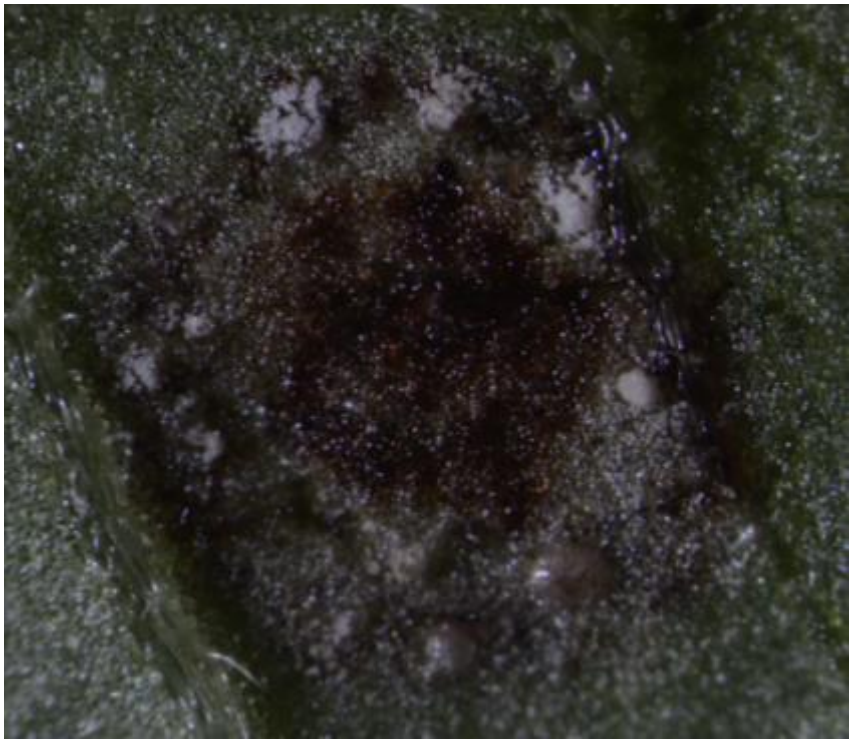
Fig. B.1. Different severity levels of soybean rust with different lesion types on different PIs of common beans. A) Big red-brown soybean rust lesions on some common bean PI. B) small tan soybean rust lesions on some common bean PIs.

(A

(B)

Fig. B.2. Red-brown lesions caused by soybean rust on common bean accessions with A) sparsely or B) no sporulating uredinia.

(A)



(B)

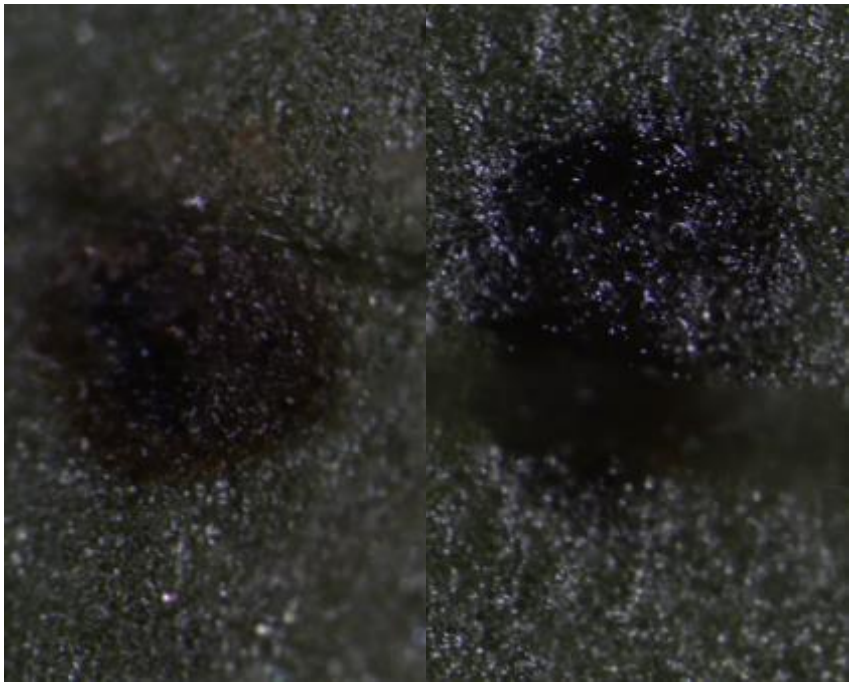
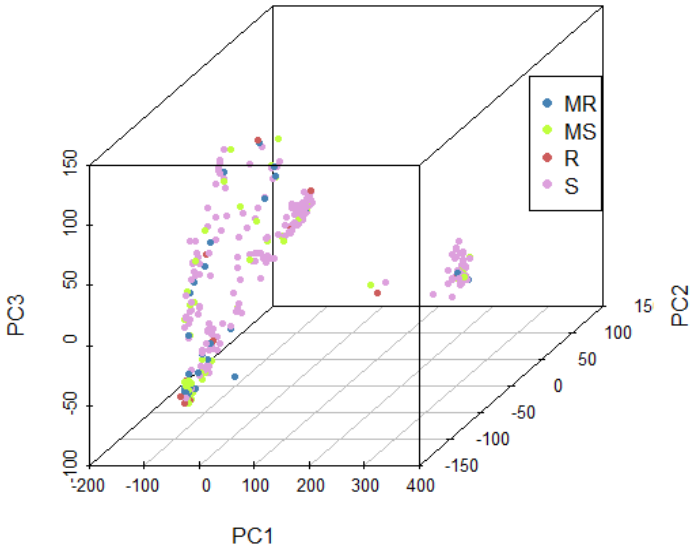


Fig. B.3. (A) Principal Component Analysis (PCA) of a diverse 363 common bean accessions genotyped with 84,416 SNPs identified through genotyping by sequencing. (B) Kinship matrix among the diverse 363 common bean accessions estimated using the 84,416 SNPs identified through genotyping by sequencing.

(A)



(B)

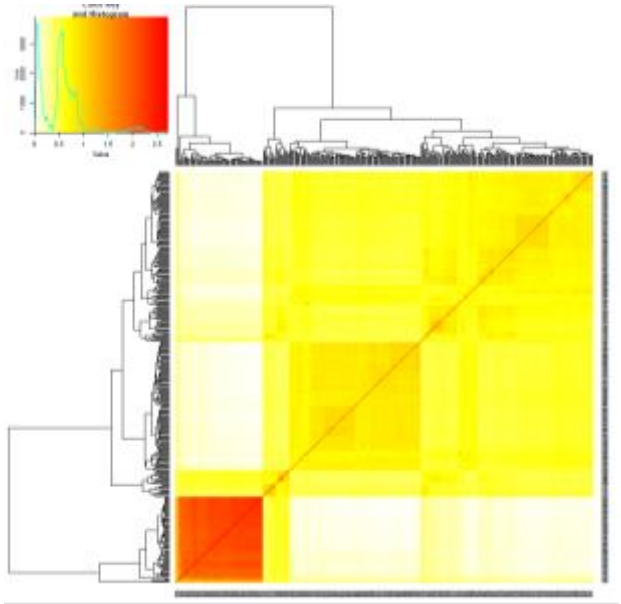
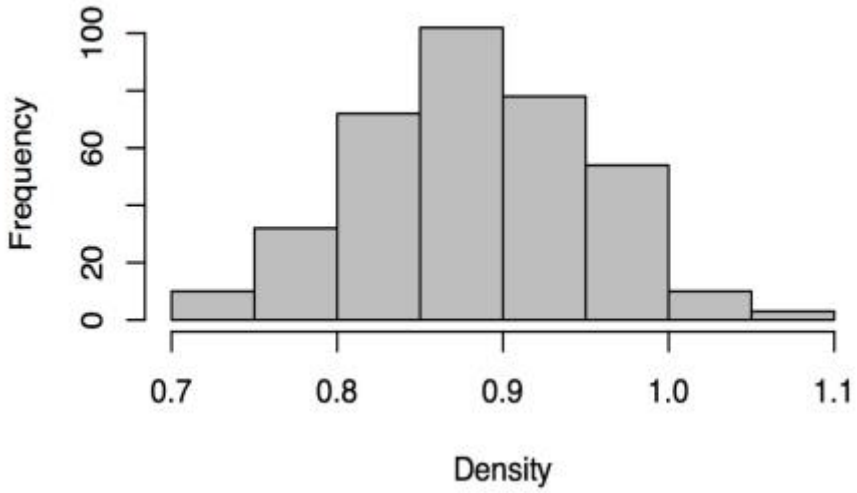


Fig. B.4. Frequency distribution of the diverse panel of 363 common bean accessions for mean soybean rust lesion numbers and Sclerotenia stem rot lesion length. Vertical axis shows the number of individuals for a given log SBR lesion number or log SSR lesion length, and the horizontal axis corresponding the log SBR lesion number and the log SSR lesion length for the 363 common bean accessions. (A) Frequency distribution of the diverse panel of 363 common bean accessions for mean soybean rust lesion numbers. (B) Frequency distribution of the diverse panel of 363 common bean accessions for mean Sclerotenia stem rot lesion length.

(A



(B)

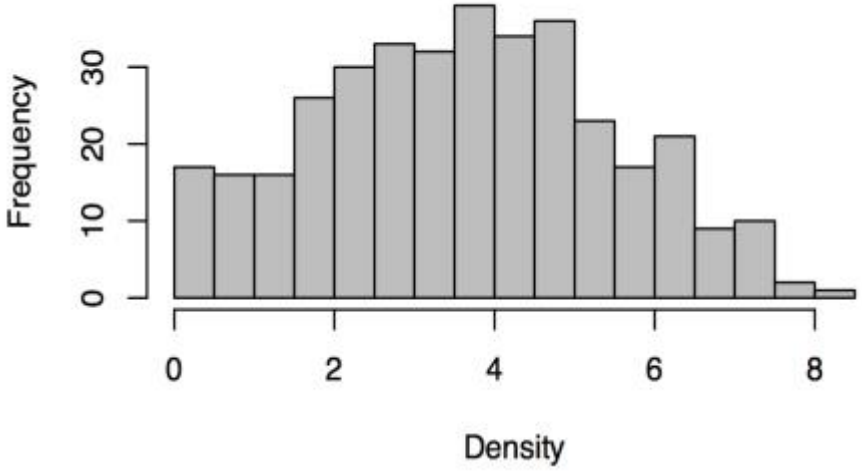


Fig. B.5 Manhattan plot produced by the genome-wide association mapping study (GWAS) showing the strength of associations of single nucleotide polymorphism (SNP) in common bean genome with soybean rust (SBR) and Sclerotenia stem rot (SSR). No SNP was detected to be significantly associated with SBR or SSR.

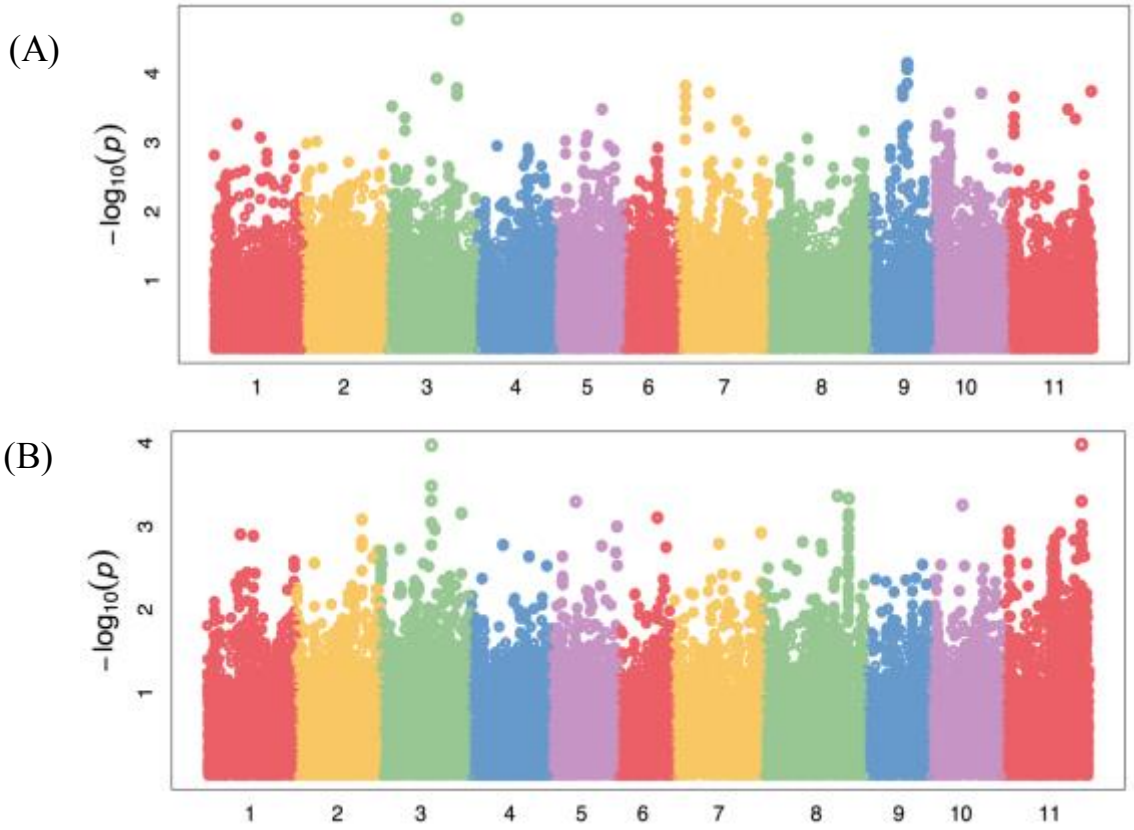
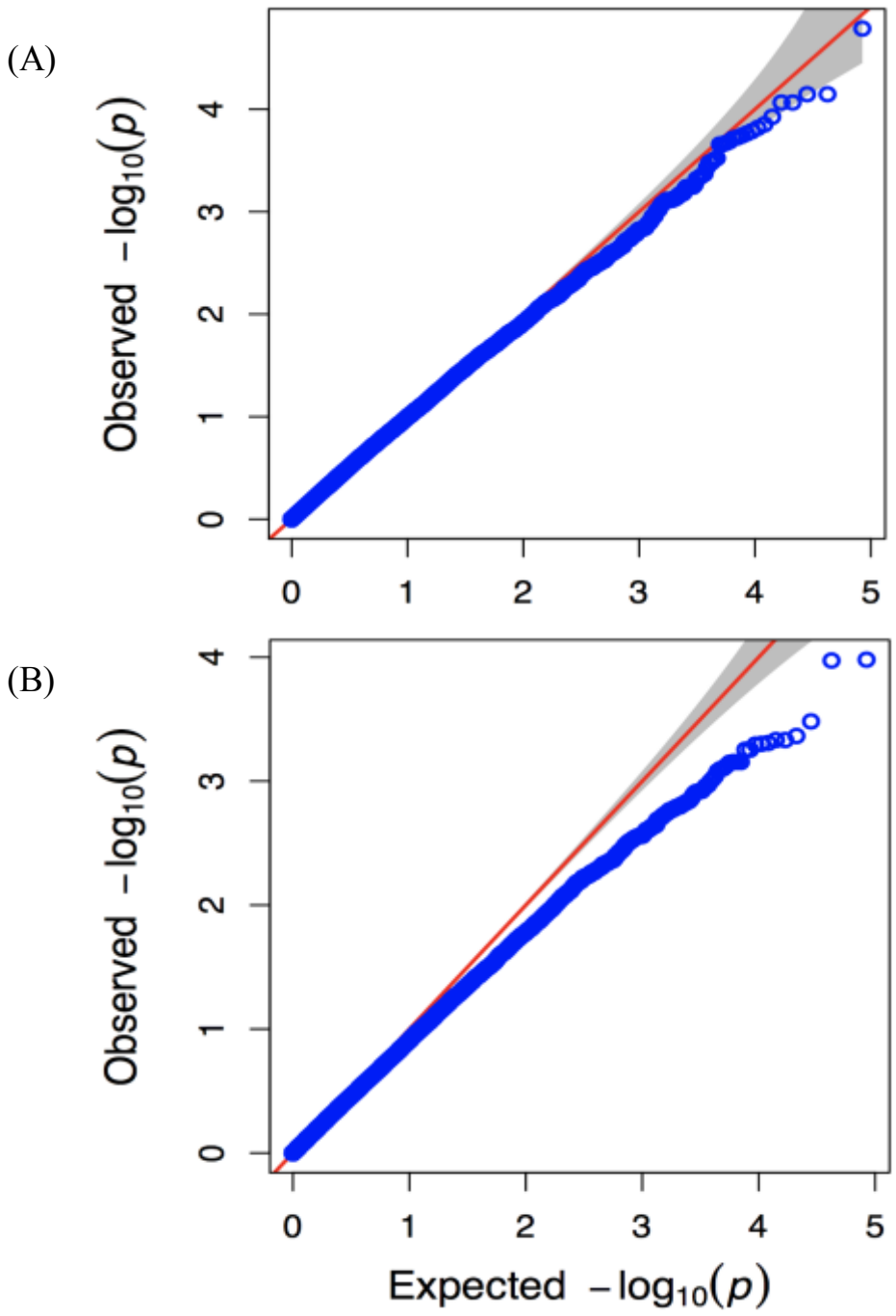


Fig. B.6. QQ plot showing quality control of the genome-wide association mapping studies (GWAS) for resistance to soybean rust (SBR) and Sclerotenia stem rot (SSR) in the core collection of common bean. (A) QQ plot of the GWAS for SBR resistance. (B) QQ plot of the GWAS for SSR resistance.



REFERENCES

Brick, M.A., Byrne, P.F., Schwartz, H.F., Ogg, J.B., Otto, K., Fall, A.L. and Gilbert, J., 2006. Reaction to three races of Fusarium wilt in the core collection. Crop Sci. 46:1245-1252.

Broughton, W.J., Hernandez, G., Blair, M., Beebe, S., Gepts, P. and Vanderleyden, J. 2003. Beans (*Phaseolus spp*.)–model food legumes. Plant and soil, 252:55-128.

Browning, B. L. and Browning, S.R. 2013. Improving the accuracy and efficiency of identity by descent detection in population data. Genetics. 194:459-71.

Chakraborty, N., Curley, J., Frederick, R. D., Hyten, D. L., Nelson, R. L., Hartman, G. L., and Diers, B.W. 2009. Mapping and confirmation of a new allele at *Rpp1* from soybean PI 504538A conferring RB lesion type resistance to soybean rust. Crop Sci. 49:783-790.

Chen, H. F., Zhao, S., Yang, Z. L., Sha, A. H., Wan, Q., Zhang, C. J., Chen, L. M., Yuan, S. L., Qiu, D. Z., Chen, S. L., Shan, Z. H., and Zhou, X. A. 2015. Genetic analysis and molecular mapping of resistance gene to Phakopsora pachyrhizi in soybean germplasm SX6907. Theor. Appl. Genet. 128:733-743.

Chung, Y. S., Sass, M.E., and Nienhuis, J. 2008. Validation of RAPD markers for white mold resistance in two snap bean populations based on field and greenhouse evaluations. Crop Sci. 48:2265–2273.

Du Preez, E.D., Van Rij, N.C., Lawrance, K.F., Miles, M.R. and Frederick, R.D. 2005. First report of soybean rust caused by Phakopsora pachyrhizi on dry beans in South Africa. Plant Dis. 89:206-206.

Garcia, A., Calvo, E. S., Kiihl, R. A. D., Harada, A., Hiromoto, D. M., and Vieira, L. G. E. 2008. Molecular mapping of soybean rust (Phakopsora pachyrhizi) resistance genes: discovery of a novel locus and alleles. Theor. Appl. Genet. 117:545-553.

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

Harris, D.K., Kendrick, M.D., King, Z.R., Pedley, K.F., Walker, D.R., Cregan, P.B., Buck, J.W., Phillips, D.V., Li, Z., and Boerma, H.R. 2015. Identification of unique genetic sources of soybean rust resistance from the USDA Soybean Germplasm Collection. Crop Sci. 55:2161-2176.

Hartman, G.L., Kull, L.S., and Huang, Y.H. 1998. Occurrence of *Sclerotinia sclerotiorum* in soybean fields in east-central Illinois and enumeration of inocula in soybean seed lots. Plant Dis. 82:560–564.

Hoffman, D.D., Hartman, G.L., Mueller, D.S., Leitz, R.A., Nickell, C.D., and Pedersen, W. L. 1998. Yield and seed quality of soybean cultivars infected with *Sclerotinia sclerotiorum.* Plant Dis. 82:826–829.

Hossain, M.M., Akamatsu, H., Morishita, M., Mori, T., Yamaoka, Y., Suenaga, K., Soares, R. M., Bogado, A.N., Ivancovich, A.J.G., and Yamanaka, N. 2014. Molecular mapping of Asian soybean rust resistance in soybean landraces PI 594767A, PI 587905 and PI 416764. Plant Path. 64:147-156.

Hyten, D.L., Hartman, G.L., Nelson, R.L., Frederick, R.D., Concibido, V.C., Narvel, J.M., and Cregan, P.B. 2007. Map location of the Rpp1 locus that confers resistance to soybean rust in soybean. Crop Sci. 47:837-840.

Hyten, D.L., Smith, J.R., Frederick, R.D., Tucker, M.L., Song, Q.J., and Cregan, P.B. 2009. Bulked segregant analysis using the GoldenGate Assay to locate the Rpp3locus that confers resistance to soybean rust in soybean. Crop Sci. 49:265-271.

Ivancovich, A. 2005. Soybean rust in Argentina. Plant Dis. 89:667-668.

King, Z.R., Harris, D.K., Pedley, K.F., Song, Q., Want, D., Wen, Z., Buck, J.W., Li, Z., and Boerma, H.R. 2015. A novel Phakopsora pachyrhizi resistance allele (Rpp) contributed by PI 567068A. Theor. Appl. Genet. doi:10.1007/s00122-015-2645-3.

Korte, A. and Farlow, A. 2013. The advantages and limitations of trait analysis with GWAS: a review. Plant Methods. 9:29.

Lipka, A.E., Tian, F., Wang, Q.S., Peiffer, J., Li, M., Bradbury, P.J., Gore, M.A., Buckler, E.S., and Zhang, Z. W. 2012. GAPIT: genome association and prediction integrated tool. Bioinformatics 28:2397-2399.

Lynch, T.N., Marois, J.J., Wright, D.L., Harmon, P.F., Harmon, C.L., Miles, M.R. and Hartman, G.L. 2006. First report of soybean rust caused by *Phakopsora pachyrhizi* on *Phaseolus* spp. in the United States. Plant Dis. 90:970-970.

McLean, R.J., and Byth, D.E. 1980. Inheritance of resistance to rust (*Phakopsora pachyrhizi*) in soybeans. Aust. J. Agric. Res. 31:951-956.

McClean, P.E., Kami, J., and Gepts, P. "Genomics and genetic diversity in common bean". *Legume crop genomics*. Wilson, R. F., Stalker, H.T., and Brummer, E.C. AOCS press, 2004. 61-82. Print.

McClean, P.E., Terpstra, J., McConnell, M., White, C., Lee, R. and Mamidi, S., 2012. Population structure and genetic differentiation among the USDA common bean (*Phaseolus vulgaris* L.) core collection. Genet. Resour. Crop Evol. 59:499-515.

Miklas, P.N., Hauf, D.C., Henson, R.A., and Grafton, K.F. 2004. Inheritance of ICA Bunsi-derived resistance to white mold in a navy × pinto bean cross. Crop Sci. 44:1584–1588.

Miklas, P.N., Porter, L.D., Kelly, J.D., and Myers, J.M. 2013. Characterization of white mold disease avoidance in common bean. Eur. J. Plant Pathol. 135:525–543.

Mkwaila, W., Terpstra, K.A., Ender, M., and Kelly, J.D. 2011. Identification of QTL for agronomic traits and resistance to white mold in wild and landrace germplasm of common bean. Plant Breed. 130:665–672.

Monteros, M.J., Missaoi, A.M., Phillips, D.V., Walker, D.R, and Boerma, H.R. 2007. Mapping and confirmation of 'Hyuuga' red-brown lesion resistance gene for Asian soybean rust. Crop Sci. 47:829-836.

Nemli, S., Asciogul, T.K., Kaya, H.B., Kahraman, A., Eşiyok, D., and Tanyolac, B. 2014. Association mapping for five agronomic traits in the common bean (*Phaseolus vulgaris* L.). J. Sci. Food Agr. 94:3141-3151.

Park, S.O., Coyne, D.P., Steadman, J.R., and Skroch, P.W. 2001. Mapping of QTL for resistance to white mold disease in common bean. Crop Sci. 41:1253–1262.

Pastor-Corrales, M.A., Sartorato, A., Liebenberg, M.M., del Peloso, M.J., Pereira, P.A., Junior, J.N. and Campos, H.D. 2007. Evaluation of common bean cultivars from the United States for their reaction to soybean rust under field conditions in Brazil and South Africa. Annu. Rep. Bean Improv. Coop. 50:123.

Pérez-Vega, E., Pascual, A., Campa, A., Giradles, R., Miklas, P.N. and Ferreira, J.J. 2012. Mapping quantitative trait loci conferring partial physiological resistance to white mold in the common bean RIL population Xana × Cornell 49242. Mol. Breed.29:31–41.

Schmutz, J., McClean, P.E., Mamidi, S., Wu, G.A., Cannon, S.B., Grimwood, J., Jenkins, J., Shu, S., Song, Q., Chavarro, C. and Torres-Torres, M., 2014. A reference genome for common bean and genome-wide analysis of dual domestications. Nature Genet. 46:707-713.

Silva, D.C.G., Yamanaka, N., Brogin, R.L., Arias, C.A.A., Nepomuceno, A.L., Di Mauro, A. O., Pereira, S.S., Nogueira, L.M., Passianotto, A. L. L., and Abdelnoor, R.V. 2008. Molecular mapping oftwo loci that confer resistance to Asian rust in soybean. Theor. Appl. Genet. 117:57-63.

Singh, S.P., Gepts, P. and Debouck, D.G. 1991. Races of Common Bean (*Phaseolus vulgaris* L., Fabaceae), Econ. Bot. 45:379-396.

Singh, S. P., and Schwartz, H.F. 2010. Breeding common bean for resistance to diseases: A review. Crop Sci. 50:2199–2223.

Shi, C., Navabi, A. and Yu, K., 2011. Association mapping of common bacterial blight resistance QTL in Ontario bean breeding populations. BMC Plant Biol. 11:1.

Soule, M., Porter, L. Medina, J., Santana, G.P., Blair, M.W., and Miklas, P. N. 2011. Comparative QTL map for white mold resistance in common bean, and characterization of partial resistance in dry bean lines VA19 and 19365–31. Crop Sci. 51:123–139.

Souza, T.L.P.O., Dessaune, S.N., Moreira, M.A. and Barros, E.G., 2014. Soybean rust resistance sources and inheritance in the common bean (*Phaseolus vulgaris* L.). Genet. Mol. Res. 13: 5626.

Vittal, R., Paul, C., Hill, C. B., and Hartman, G. L. 2014. Characterization and quantification of fungal colonization of *Phakopsora pachyrhizi* in soybean genotypes. Phytopathology. 104:86-94.

Vuong, T.D., D.D. Hoff man, B.W. Diers, J.F. Miller, J.R. Steadman, and G.L. Hartman. 2004. Evaluation of soybean, dry bean, and sunflower for resistance to *Sclerotnia sclerotiorum*. Crop Sci. 44:777–783.

Yorinori, J.T, Paiva, W.M., Frederick, R.D., and Costamilan, L.M. (2005). Epidemics of soybean rust (*Phakopsora pachyrhizi*) in Brazil and Paraguay from 2001 to 2003. Plant Dis. 89: 675-677.

Zhang, Z., Ersoz, E., Lai, C.Q., Todhunter, R.J., Tiwari, H.K., Gore, M.A., Bradbury, P.J., Yu, J., Arnett, D.K., Ordovas, J.M., and Buckler, E.S. 2010. Mixed linear model approach adapted for genome-wide association studies. Nat. Genet. 42:355–360.