CHARACTERIZING *Rhg1* MEDIATED SOYBEAN RESISTANCE TO SOYBEAN CYST NEMATODE USING FUNCTIONAL AND COMPUTATIONAL APPROACHES

By

David E. Cook

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Plant Pathology)

at the

UNIVERSITY OF WISCONSIN-MADISON

2013

Date of final oral examination: 08/06/13

The dissertation is approved by the following members of the Final Oral Committee:

Andrew F. Bent, Professor, Plant Pathology

Ann MacGuidwin, Professor, Plant Pathology

Jeri Barak, Assistant Professor, Plant Pathology

Michael Havey, Professor, Horticulture

Jean-Michel Ane, Associate Professor, Agronomy

CHARACTERIZING *Rhg1* MEDIATED SOYBEAN RESISTANCE TO SOYBEAN CYST NEMATODE USING FUNCTIONAL AND COMPUTATIONAL APPROACHES

David E. Cook

Under the supervision of Professor Andrew F. Bent

At the University of Wisconsin-Madison

ABSTRACT

Domesticated soybean, *Glycine max*, is an important world commodity accounting for 68% of world protein meal and 56% of world oilseed production. Soybean cyst nematode (*Heterodera glycines*) causes billions of dollars of economic losses annually and is believed to be the most economically damaging disease of soybean. Significant research has been devoted to identifying natural genetic variation for soybean cyst nematode resistance. A number of loci have been identified in *Glycine max* that contribute to resistance, but a single locus termed *Rhg1* has been identified repeatedly as having the greatest impact on resistance and is currently deployed in most commercially utilized soybean cultivars that are designated as resistant to soybean cyst nematode. The mechanism underlying *Rhg1*-mediated soybean cyst nematode resistance has remained elusive and was the topic of my research.

This dissertation focuses on functional and computational approaches to define and characterize the genes underlying *Rhg1*-mediated soybean cyst

nematode resistance. First, previous claims that a leucine-rich repeat kinase protein controls *Rhg1*-mediated resistance were tested. Using functional genetics along with pathogen bioassays, we were unable to detect significant impacts on nematode resistance when silencing or strongly expressing the leucine-rich repeat kinase. To identify the genes responsible for *Rhg1*-mediated resistance, we used similar assays to test for functional impacts of genes within the recently determined fine structure genetic map interval for *Rhg1*. Unexpectedly, we identified three tightly clustered genes at the *Rhg1* locus that each contribute to *Rhg1* mediated soybean cyst nematode resistance. We further discovered that the DNA encoding these genes is present in multiple copies in soybean cyst nematode-resistant parents, and this causes elevated expression of the genes. Two of the identified genes, *Glyma18g02580* and *Glyma18g02610*, did not carry amino acid polymorphisms between resistant and susceptible *Rhg1* haplotypes. The third gene, *Glyma18g02590*, encoding a predicted α-SNAP protein did contain amino acid polymorphisms relative to the reference soybean genome, which is based on the soybean cyst nematode-susceptible line Williams 82. Transgenic expression of any one of these three genes in soybean roots did not discernibly improve resistance, but simultaneous expression of all three genes did enhance resistance to soybean cyst nematodes. To further explore the evolution and diversity of *Rhg1,* we used whole genome sequencing, fiber-FISH and related techniques to examine the structural and nucleic acid variation in the genomes of four-dozen soybean accessions. We discovered that the *Rhg1* locus is commonly arranged in a single

copy in all tested SCN susceptible germplasm, but is present as either a low-copy or high-copy type in SCN-resistant germplasm. These repeat copy classes are also distinguishable by expression level differences and the presence of related but distinct alleles of the previously identified α-SNAP protein. We also explored the role of DNA-methylation differences at the locus and their implication in modulating disease resistance. The identification of the genes that control *Rhg1*-mediated resistance, and an understanding of their evolution and diversity, should foster efforts to improve the disease resistance that is available to reduce the deleterious impacts of soybean cyst nematodes.

ACKNOWLEDGEMENTS

None of this would have been possible without the love, support, and laughs from my two daughters, Lucile and Nellie, and the love of my super human wife, Annie. I also want to thank my parents, Dawn and Peter for their years of support and encouragement, especially when I did not deserve it. And to my sister, Katie, for always being available to listen and offer a kind word. Thank you to Andrew F. Bent, for fostering and allowing me to grow as a scientist. You have been a great mentor, and I will never forget your influence. I have worked with a number of great researchers in the Bent Lab, and I would like to specifically thank Xiaoli Guo for her collaboration and fun conversations about life and nematodes, Sara Melito for teaching me so many useful skills, and Adam Bayless for continued innovation, stimulating conversation and coffee. I am also thankful for the time Laura Helft and I spent across the lab bench engaged in serious and seriously ridiculous conversation. Thank you to my thesis committee members Jeri Barak, Ann MacGuidwin, Jean-Michel Ane, and Michael Havey for providing expert advice. Thank you also to the Department of Plant Pathology for offering a diverse and exciting atmosphere to conduct research.

TABLE OF CONTENTS

1. Introduction

1.1. Soybean (Glycine max)

Domesticated soybean (*Glyine max*) is a worldwide source of oil and protein, and vitally important for a diverse set of industries. Soybean accounts for the second most planted crop acreage in the United States, an estimated 77.7 million acres for 2013 and is the second most valuable crop with a reported value for 2012 of $43.2 billion (soystat.com). This puts soybean's value at more than half the value of corn ($77.4 billion) and significantly more valuable than the next row crop, wheat ($17.9 billion) (NASS, 2012). Soybeans produced in the United States are used domestically and globally; China is the number one importer of whole soybeans, Canada is the number one importer of soybean meal, and Morocco is the number one importer of soybean oil (soystat.com). While many countries continue to debate the use and safety of genetically modified crops, 94% of the soybeans grown in the United States in 2011 were registered as genetically modified (soystat.com).

The release of the *Glycine max* genome sequence and annotation in 2010, detailed the 1,115 megabases (Mb) of DNA, of which approximately 950 Mb are arranged on anchored assemblies (Schmutz et al., 2010) . The sequence was developed from the soybean variety Williams 82, and is estimated to encode >46 thousand proteins. Evidence suggests the soybean genome has undergone two rounds of whole genome duplication, the first having occurred roughly 59 million years ago with a second, more recent duplication occurring 13 million years ago (Doyle et al., 2003; Innes et al., 2008; Schmutz et al., 2010). The original whole genome duplication is thought to

have occurred within the papilionoid (also called *Faboideae*) subfamily of *Fabaceae* (Doyle and Luckow, 2003). The more recent genome duplication occurred within the *Glycine* genus, resulting in many loci having homologuos sequence on at least one other chromosome. Soybean is referred to as a paleopolyploid, but functions as a diploid with 20 chromosomes.

1.2. Soybean Cyst Nematode (*Heterodera glycines*)

The most economically damaging pathogen of soybean is the soybean cyst nematode (SCN, *Heterodera glycines*). Of the total soybean yield losses in the U.S. caused by the top 28 pathogens, soybean cyst nematode accounted for 20% – 37% of the total losses estimated from 2005 to 2010 (Wrather and Koenning, 2006, 2009). Soybean cyst nematode was first discovered in North Carolina in the 1950s, and a survey in 2006 found SCN in every major soybean producing state (Niblack, 2005).

Soybean cyst nematode is an obligate endoparasitic animal, infecting the roots of a limited number of plant species (Lauritis et al., 1983; Cloud et al., 1988; Niblack et al., 2004; reviewed in Niblack et al., 2006). Soybean cyst nematode overwinters in soil as a protective cyst, housing hundreds of eggs. SCN molts through a number of life stages each with a characteristic morphology, before reaching reproductive maturity. The first molt to the juvenile 1 (J1) stage takes place within the egg,

followed by a subsequent molt to the J2 stage, which is the mobile, infective stage of SCN. Following root penetration at the zone of elongation, J2 nematodes travel intracellularly through the cortex to the vasculature using both mechanical force and secreted proteins, termed effectors (Smant et al., 1998; Davis et al., 2004). A single cell, often a pericycle cell adjacent to the vasculature, is selected to establish and maintain the feeding site, or syncytium. The selected feeding cell undergoes a number of cellular changes including increased cytoplasmic streaming, increased ribosome activity, vacuole enlargement, increased cell wall thickening and invagination, and surrounding cells are incorporated via cell wall dissolution (Endo, 1984; Williamson and Hussey, 1996; Davis et al., 2004). This is a highly evolved and amazing biological phenomena, involving cellular re-differentiation of plant roots by a nematode, orchestrated by secreted nematode effectors (Gheysen and Mitchum, 2011; Mitchum et al., 2012). This initial phase of syncytial development is similar in roots of both resistance and susceptible soybeans (Li et al., 2004). Once a syncytium is established, the nematode is sedentary and undergoes additional molts. Males molt to the J3 stage, and following a final molt emerge as vermiform nematodes, capable of traveling in search of females to fertilize. Females reach reproductive maturity following molts to the J3 and J4 stage, which have the salient swollen appearance similar to a lemon. After fertilization, the female continues to feed, never moving from the initial syncytium. Eggs develop both within the body cavity, which will serve as the following years inoculum, and outside the body for additional intra-season infection. The female nematode eventually dies and her

body melanizes, becoming the protective cyst for overwintering. Soybean cyst nematode is particularly problematic because once present in a field, it cannot feasibly be eradicated, and pathogenic SCN eggs can overwinter in cyst for greater than a decade (Niblack et al., 2006).

Current control strategies focus on reducing SCN inoculum level through a combination of planting SCN resistance soybean varieties and crop rotation with non-hosts (Williamson and Hussey, 1996; Niblack, 2005; Kulkarni et al., 2008) Non-host crops include corn, wheat, tomato, alfalfa, sugar beet, cotton, peanut and many others, but the economic reality of farming often dictates a continual corn- soybean rotation. Yearly planting of soybean in SCN infested fields creates a strong selective pressure for SCN populations to overcome deployed resistant varieties, a scenario already noted (Niblack et al., 2008). This potential has also been reported in laboratory settings (Colgrove and Niblack, 2008).

1.3. Plant Defense to Cyst Nematodes

Cyst forming plant parasitic nematodes are not only a major problem on soybean, but also cause significant yield losses on potato, tomato, sugar beet, wheat, and rice. Two types of cyst nematodes*, Globodera rostochiensis* and *Globodera pallida*, infect potato and tomato (Sobczak et al., 2005)*; Heterodera schachtii* infects sugar beet (Cai et al., 1997); *Heterodera avenae* infects wheat, oats, barley, and rye; *Heterodera*

oryzae infects rice. This is not a comprehensive list of *Heterodera* species or of the cyst nematodes infecting a given crop, but rather a list of common crop - cyst nematode pairs. For all crops, cyst nematodes can cause significant yield losses if inoculum levels are high, and they cannot easily be exterminated once present in a field. In all cases, breeding programs focus on developing crops with enhanced host resistance.

While a number of resistance loci have been mapped from various crops, relatively few genes have been cloned. A summary of cloned resistance genes to cyst nematodes include, *Hero* cloned from tomato encodes a *CC-NBS-LRR* type R-gene (Ernst et al., 2002), and *Gro1-4* and *Gpa2,* each isolated from potato, encode *TIR-NBS-LRR* type R-genes (van der Vossen et al., 2000; Paal et al., 2004). The $Hs1^{pro-1}$ resistance gene cloned from the wild sugar beet, *Beta procumbens*, is predicted to contain a leucine rich region and a transmembrane domain (Cai et al., 1997), which does not strictly fit into one of the plant resistance genes classes (Hammond-Kosack and Jones, 1997). An additionally important but not cloned resistance locus, termed *H1* in potato, has been mapped to a small interval, which contains a NBS-LRR gene cluster (Finkers-Tomczak et al., 2011). Taken together, the current state of known resistance proteins to plant parasitic nematodes either have the same or homologous structure to canonical plant immune receptors (Bent, 1996; Ellis et al., 2000).

The most commonly used and initially developed sources of soybean resistance to SCN are derived from lines Peking, PI 88788, PI 90763, PI 437654, PI 209332, PI 89772, and Cloud- lines which are now used to determine the virulence type of nematode field populations (Kim et al., 1987; Diers et al., 1997; Niblack et al., 2002). These lines were selected based on their early adoption as breeding material for SCN resistance, not necessarily because they have different SCN resistance reactions. In fact, it was recently determined that the resistance response, measured as nematode development, is significantly correlated for lines PI 88788, PI 209332, and Cloud, while Peking, PI 90763, and PI 89772 form a second correlated resistance group (Diers et al., 1997; Colgrove and Niblack, 2008). The resistance response of PI 437654 was similar to the Peking group, but gave variable results depending on the nematode inoculum. This suggests a narrow genetic base breeding material commonly used to develop SCN resistance varieties. Interestingly, while PI 88788 and Peking are considered effective to a broad range of nematode populations, their respective resistance responses are not identical at the cellular level (Mahalingam and Skorupska, 1996; Kim et al., 1998; Kim et al., 2010). The Peking type resistance has been characterized as rapid with a pronounced cell death surrounding the nematode head (Endo, 1978; Mahalingam and Skorupska, 1996). Contrasting this, PI 88788 resistance is characterized by its slower response lacking signs of obvious cell death and cell wall appositions (Matsye et al., 2011). Despite the reported more subtle cellular response, PI 88788 is the primary source of commercial SCN resistance, and significantly reduces nematode populations in the

field (Concibido et al., 2004; Brucker et al., 2005). Many researchers now recognize PI 88788 and Peking-type of SCN resistance, while PI 437654 resistance resembles aspects of both responses (Anand, 1992; Colgrove and Niblack, 2008) .

Significant effort has been placed on identifying a broader base of SCN resistant germplasm (Raoarelli et al., 1992; Webb et al., 1995; Arelli and Webb, 1996; Kim et al., 1998). Screening the USDA soybean germplasm collection identified 118 plant introductions (PI) showing some level of SCN resistance (Arelli et al., 2000). Still, few of these lines are used for commercial cultivar development owing to the common problems linkage drag of undesirable traits during traditional breeding with less adapted material. Taken together, numerous mapping efforts to identify markers linked to SCN resistance genes have yielded inconsistent results (Concibido et al., 2004). A total of 61 quantitative trait loci (QTL) had been identified in nearly 20 studies, but many were only found in a single study and display a minor effect on resistance. There is however, a single QTL on chromosome 18, termed *Rhg1* (*resistance to Heterodera glycines*), which has been repeatedly identified and accounts for a major portion of the resistance phenotype (Cregan et al., 1999; Glover et al., 2004; Brucker et al., 2005). The other commonly identified resistance gene is *Rhg4*, but its contribution to SCN resistance is only measured in certain soybean lines (Webb et al., 1995; Wu et al., 2009).

Resistance derived from PI 88788, termed *Rhg1-b*, is present in most (>90%) of the commercially available SCN resistant soybean cultivars in the U.S. While the locus was identified greater than 50 years ago (Caldwell et al., 1960), it still provides a high level of resistance. In the last 20 years, the development of molecular markers linked to *Rhg1* has assisted SCN resistance breeding efforts (Webb et al., 1995; Cregan et al., 1999), and marker assisted selection is now widely used in commercial breeding programs to select for *Rhg1*.

Many laboratories have studied the genetic and molecular basis of *Rhg1* over the past 20 years. One study of note identified a gene encoding a Leucine Rich Repeat-Kinase (LRR-Kinase) at the *Rhg1* locus (Ruben et al., 2006). LRR-kinases are a large family of protein kinases involved in a diverse role of cellular processes, one being plant defense. The researchers proposed this LRR-kinase, residing in proximity to *Rhg1* markers, was the gene controlling *Rhg1* mediated SCN resistance - patent applications were also filed (Lightfoot and Meksem, 2002). Monsanto researchers proposed a similar hypothesis in separate patent applications (Hauge et al., 2001). To date, no conclusive research has been published demonstrating the *Rhg1* LRR-kinase impacts on SCN development (Melito et al., 2010), and patent claims to this effect have not been granted. More recent fine-structure mapping of *Rhg1-b* identified crossover events that further delimited the genetic locus to a 67 kilobase (kb) interval, predicted to encode 11 genes based on the reference genome Williams 82, a SCN susceptible variety (Kim et al., 2010).

A recent study describes the impact of an isoform of an α-SNAP gene residing within the *Rhg1* locus (Matsye et al., 2012). The researchers identified multiple SNPs in both Peking and PI 437654, not present in the reference Williams 82 or PI 88788. One SNP is predicted to cause read through of an exon-intron boundary, resulting in an additional 17 transcribed nucleotides before encountering an in-frame stop codon (Matsye et al., 2012). The isoform was confirmed via cDNA sequencing, but evidence is not provided that the resulting mRNA is translated. Overexpressing the cDNA variant in the susceptible genotype Williams 82 resulted in plants with enhanced resistant to SCN. These experiments were carried out in chimeric plants, which have a native above ground portion and transgenic roots. Constitutive expression of several well-defined defense responsive genes such as *GmEREBP1, PR1, PR2,* and *PR5* was however reported in transgenic compared to control roots. Expression of these genes was previously reported to reduce nematode development in *Arabidopsis*, confounding the observed increased resistance. Another paradox with this report is the observation that PI 88788, the most widely used source of commercial SCN resistance, contains the same allele as Williams 82. The authors acknowledge this, and suggest the altered allele is responsible not for resistance *per se*, but affects cell wall appositions (CWA) present in roots of Peking and PI 437654 undergoing a resistance response that are not present in roots of PI 88788 (Endo, 1978). This is an interesting hypothesis that requires further investigation. There is, however, one concern with the interpretation of the

identified α-SNAP (Matsye et al., 2012). Closer inspection of the reported SNPs indicates the mRNA isoform of α-SNAP is encoded by a paralog on chromosome 11, not by the α-SNAP gene at *Rhg1* chromosome 18. This claim is based on sequence alignments and confirmed by whole genome sequence analysis of 14 SCN resistant lines, data for which is presented in chapter 4 of this thesis.

Unlike the PI 88788 source of resistance, Peking-type resistance requires another locus, *Rhg4*, for full resistance (Meksem et al., 2001; Brucker et al., 2005). In a hallmark study, the effect of different *Rhg4* alleles in different *Rhg1* backgrounds was measured in recombinant inbred lines (RIL) from a cross of PI 88788 and PI 437654 type resistance (Brucker et al., 2005). The authors showed that in RILs containing *Rhg1* from PI 88788, having an *Rhg4* allele from PI 43765 vs PI 88788 results in a consistent but small increase in resistance. By contrast, RILs containing *Rhg1* from PI 437654, having an *Rhg4* allele from PI 437654 vs PI 88788 resulted in a roughly 500% increase in resistance (Brucker et al., 2005). These results suggest two important conclusions. The *Rhg1* alleles from PI 88788 and PI 437654, determined by genetic markers, are different, and *Rhg4* functions in conjunction with either *Rhg1* allele, but its effect on resistance is more pronounced in a PI 437654 *Rhg1* background. It is not clear, however, if the function of PI 437654-*Rhg4* is modified by different *Rhg1* alleles, or if *Rhg1* from PI 88788 is so effective that it is difficult to detect *Rhg4* in this background. Further discussion about the

different levels of resistance and roles of *Rhg1* and *Rhg4* are presented in chapter 4 of this thesis.

The gene controlling *Rhg4* mediated resistance was recently cloned and found to be a serine hydroxymethyltransferase (SHMT) (Liu et al., 2012). The evidence for this is based on genetic mapping, sequencing, plant mutants, *in vitro* assays, and biochemical analysis. Analysis of SCN resistance in mutant lines showed that the normally resistant plant line with heterozygous *Rhg4* alleles allows roughly 4 times more nematode development than the homozygous Rhg4 resistant lines. Also, the homozygous mutant supported almost 8 times more nematode development, going from a female index of 5 to almost 40. The identification of intermediate resistance in the heterozygous and homozygous lines suggests incomplete dominant gene action. It is worth noting that in the homozygous *Rhg4* mutants, nematode development is still only 38% of that seen on the susceptible control. This result again illustrates the utility of *Rhg1* mediated resistance, which likely accounts for a great proportion of the nearly 62% decrease in nematode development on these lines.

Why a SHMT would be involved in soybean resistance to SCN remains an intriguing question. The canonical cellular function of SHMT is in one carbon metabolism, where it is responsible for the reversible conversion of serine to glycine through the removal of serine's methyl group with THF being the main acceptor (Hanson and Roje, 2001). The reduced rate of serine to glycine conversion by the

resistant allele of SHMT shown in (Liu et al., 2012) does raise the intriguing possibility that reducing the amount or rate of one carbon metabolism may negatively affect nematode development. There are alternative hypothesis for how *Rhg4* affects nematode development, and additional studies are needed.

In addition to genetic mapping, a number of studies have analyzed soybean's response to SCN infection using microarray technology to quantify differences in mRNA abundance between samples (Alkharouf et al., 2006; Klink et al., 2007, 2007; Klink et al., 2010; Kandoth et al., 2011; Klink et al., 2011; Matsye et al., 2011). These studies provided the first insights into the broad class of cellular responses being evoked in different SCN resistant soybean lines, but also lack detailed resolution due to technologic limitations of their time. One drawback of published soybean-SCN microarrays studies is the use of the GeneChip Soybean Genome Array from Affymetric, which only contains probes for roughly half of soybean's predicted open reading frames, and fails to distinguish many paralogs (Libault et al., 2010; Valdes-Lopez et al., 2011). Another limitation of most published microarrays on plant-nematode interactions is that they extract RNA from entire infected roots. This is problematic for SCN infection experiments because of the exact spatial nature of infection, That is, in any given infected root, only a few cells are contacted and directly impacted by nematode infection. Therefore, the number of infected cells collected from an entire root would be several orders of magnitude lower than the

number of non-infected root cells, which can results in a significant noise to signal ratio.

Two recent microarrays utilized laser-capture microdissection to collect tissue from specific cell types including syncytium and surrounding cells (Kandoth et al., 2011; Matsye et al., 2011). One study compared the resistance response in near isogenic lines with SCN resistance derived from PI 209332 (Kandoth et al., 2011), while the other study examined expression differences in different cell types of Peking and PI 88788 (Matsye et al., 2011). Of the 1,447 genes differentially regulated in Kandoth et al. 2011, 17% are predicted to be involved in stress and defense, with a number of cell death, salicylic acid, and jasmonic acid responsive genes being up regulated during defense. Many of the same pathways are identified in Matsye et al. 2011, but they also note the enrichment of genes expressed in pathways relating cell wall re-enforcement and vesicle trafficking.

Matsye et al. 2011, also specifically looked at genes residing in the *Rhg1* genetic interval (Kim et al., 2010) and identified that *Glyma18g02580* and *Glyma18g02590* are specifically expressed in syncytium undergoing a defense response. They mention that no other probe sets exist for other genes in the *Rhg1* interval (9 other genes), but some of their data indicate otherwise. In general, there are a number of inconsistencies in the paper and the lack of publicly available data makes interpretation difficult. Manual inspection of the data provided by Kandoth et al. 2011 for genes *Glyma18g02580, Glyma18g02590,* and *Glyma18g02600* indicates

these genes are 4, 4, and 1.5 fold respectively more highly expressed in resistant syncytium at 5 and 8 days post inoculation. No probe sets exist for *Glyma18g02610.* These genes are specifically mentioned here in light of findings presented in chapter three.

1.4. Structural Variation and Genome Evolution

There is now considerable evidence and growing consensus that genome structural variation, defined as insertions, deletions, and inversions of variable size (Feuk et al., 2006) are prevalent and contribute to phenotypic variation in eukaryotes at least on the same order as SNPs (Perry et al., 2008; Stankiewicz and Lupski, 2010; Girirajan et al., 2011). Their contribution took longer to realize, owing mainly to technical limitations to identify, characterize, and quantify structural variation present in a genome. Adoption of hybridization-array techniques and massive advances in whole genome sequencing and computation has opened the door to study this type of genetic variant (Alkan et al., 2011).

A specific type of structural variation, termed copy number variation (CVN), is pertinent and a focus of this thesis. CNV is defined here as a segment of homologous DNA occurring more than once in a genome (Aitman et al., 2006; Conrad and Antonarakis, 2007). Copy number variant segments can range from 50bp to 1Mb or larger, and be arranged as tandem duplications or be interspersed within or

between chromosomes. There is considerable evidence that copy number variation can affect transcription (Stranger et al., 2007; Schlattl et al., 2011), and is related to a number of adaptive traits (Brown et al., 1998; Labbe et al., 2007; Perry et al., 2007)

The evolutionary forces affecting copy number variants can be viewed in the framework of gene duplication (Lynch and Conery, 2000; Innan and Kondrashov, 2010; Kondrashov, 2012). Gene duplication arises from either an unequal exchange during recombination or from the action of transposable nucleic acids. For a comprehensive review on mechanisms of gene duplication see (Hastings et al., 2009). Gene duplicates or regions of duplicates create copy number variants, but a number of possible outcomes can ensue for the duplicates. The most frequent outcome is the accumulation of polymorphisms, resulting in the loss of function or expression, referred to as pseudogenation (Lynch and Conery, 2000). The other broad class of possibility is the duplicate will acquire beneficial or deleterious mutations and be driven to a high or low frequency in the population (Kondrashov and Kondrashov, 2006; Innan and Kondrashov, 2010). While most random change will perturb an organism's previously evolved physiology, some mutations can be beneficial either because of increased gene dosage or the acquisition of a novel function (Conant and Wolfe, 2008). Note, gene duplication can be viewed in the context of whole genome duplications, which is not considered an example of copy

number variation, and is a different evolutionary event leading to paralogous genes on separate chromosomes.

There are several examples of copy number variation playing a significant role in the evolution of organisms spanning most kingdoms of life (Segovia, 1994; Perry et al., 2008; Iskow et al., 2012; Maron et al., 2013). One report identified a large expansion in an antifreeze glycoproteins gene in Antarctic cod, relative to warm water species, which the authors associate with adaptation to life in cold waters (Chen et al., 2008). Numerous examples of bacterial adaptation to antibiotics, in which copy number increases for transporters and efflux pumps lead to increased tolerance, have been reported from both natural and laboratory settings (Kondrashov and Kondrashov, 2006; Craven and Neidle, 2007). COULD ADD A COUPLE (3 MORE EXAMPLES to bump up).

The occurrence of copy number variation affecting traits in plants is particularly prevalent, likely owing to their relatively large, repetitive, plastic genomes (Moore and Purugganan, 2005; Yu et al., 2005; Flagel and Wendel, 2009). Adaptation to environmental stress via copy number increase has emerged as a common theme in plant genome evolution. Extensive work on comparative genomics of model plants has found interesting links between the life histories and genome content of *Arabidopsis thaliana* and *Thellungiella parvula* (Dassanayake et al., 2011). The authors identified 1,278 tandem duplications for genes in *T. parvula,* compared to

1,113 tandem duplications in *A. thaliana*, but only half were shared between the genomes. Interestingly, the authors grouped the different duplicated genes based on predicted function (gene ontology, GO) and found that *T. parvula* had statistically significantly more copy number variant genes in the categories of transport, response to abiotic/biotic stimulus, and developmental processes. This is interesting based on the fact that *T. parvula* naturally occurs in areas of high salt content and low nutrient availability, suggesting a mechanism for adaptation to extreme environmental conditions in *T.parvula* is through copy number increases of relevant genes (Oh et al., 2012).

Recent work showing adaptation to abiotic stress for a specific trait highlights the importance of gene duplication and copy number variation in plant evolution. In *Zea maize*, increased copy number of a specific transporter in the multidrug and toxic compound extrusion (MATE) family was identified as explaining the quantitative differences in aluminum tolerance in a RIL population (Maron et al., 2013). The encoded transporter does not contain polymorphic amino acids between aluminum tolerant and sensitive plants, but the lines with high aluminum tolerance contain more copies of the gene. A near isogenic line of the aluminum sensitive parent with the addition of the three copies of *MATE1* was roughly 2 times less sensitive to aluminum, indicating the addition of the three copies alone accounts for a significant portion of aluminum tolerance.

1.5. Justification and Goals

Despite the significant amount of work and importance of SCN resistance, the genetic and molecular mechanisms of soybean resistance to soybean cyst nematode have remained unclear. A number of resistance loci have been reported in soybean, but few show consistent results and none had been cloned. Results from microarray experiments are useful for elucidating broad pathways associated with cyst nematode resistance, but again lack detail about the genes and ultimately proteins responsible for SCN resistance. There are also instances where the published literature for *Rhg1* mediated resistance appears inaccurate.

One such point of contention is the role the LRR-Kinase situated adjacent to the *Rhg1* interval plays in SCN resistance. My initial goal was to assist Sara Melito and Adam Hueburger in experiments they started, directly testing the effect of silencing the Rhg1 LRR-kinase on nematode development (Melito et al., 2010). Data for this project is published and presented in chapter two of this thesis.

Following the work presented in chapter two, the goal was to use a candidate gene silencing approach to identify which genes at *Rhg1* contribute SCN resistance. This work identified 3 separate genes at *Rhg1* each contributing to SCN resistance, and has opened many avenues of continued research (Cook et al., 2012). Details of these findings, including a collaborative effort that led to the discovery of copy number variation of *Rhg1* are published and presented in chapter three.

The identification of copy number variation at *Rhg1* raised a number of questions regarding differences seen across SCN resistant germplasm. The goal of the work presented in chapter four is to characterize the copy number variation present in a population of soybean for absolute copy number, gene expression, structure, content, evolution, and DNA methylation. Knowing more about variation at the locus will provide new goals and methods for breeding SCN resistance soybean varieties, potentially provide valuable insight into the mechanism of resistance and be one of the most detailed analyses of a copy number locus in any organism. Data presented in chapter four are being prepared for submission at this time.

2. A nematode demographics assay in transgenic roots reveals no significant impacts of the Rhg1 locus LRR-Kinase on soybean cyst nematode resistance

2.1.

This chapter was previously published as:

Melito S., Heuberger A., Cook D.E., Diers B., MacGuidwin A., Bent A.F. 2010. A nematode demographics assay in transgenic roots reveals no significant impact of the Rhg1 locus LRR-Kinase on soybean cyst nematode resistance. BMC Plant Bio 10:104

I assisted in replicating the transgenic complementation assay results presented in the publication, along with assisting in writing a portion of the manuscript.

2.2. Abstract

Background: Soybean cyst nematode (*Heterodera glycines,* SCN) is the most economically damaging pathogen of soybean (*Glycine max*) in the U.S. The *Rhg1* locus is repeatedly observed as the quantitative trait locus with the greatest impact on SCN resistance. The Glyma18g02680.1 gene at the *Rhg1* locus that encodes an apparent leucine-rich repeat transmembrane receptor-kinase (LRR-kinase) has been proposed to be the SCN resistance gene, but its function has not been confirmed. Generation of fertile transgenic soybean lines is difficult but methods have been published that test SCN resistance in transgenic roots generated with *Agrobacterium rhizogenes*.

Results: We report use of artificial microRNA (amiRNA) for gene silencing in soybean, refinements to transgenic root SCN resistance assays, and functional tests of the *Rhg1* locus LRR-kinase gene. A nematode demographics assay monitored infecting nematode populations for their progress through developmental stages two weeks after inoculation, as a metric for SCN resistance. Significant differences were observed between resistant and susceptible control genotypes. Introduction of the *Rhg1* locus LRR-kinase gene (genomic promoter/coding region/terminator; Peking/PI 437654-derived SCN-resistant source), into *rhg1*$^{-}$ SCN-susceptible plant lines carrying the resistant-source *Rhg4*$^{+}$ locus, provided no significant increases in SCN resistance. Use of amiRNA to reduce expression of the LRR-kinase gene from the *Rhg1* locus of Fayette (PI 88788 source of *Rhg1*) also did not detectably alter

resistance to SCN. However, silencing of the LRR-kinase gene did have impacts on root development.

Conclusion: The nematode demographics assay can expedite testing of transgenic roots for SCN resistance. amiRNAs and the pSM103 vector that drives interchangeable amiRNA constructs through a soybean polyubiqutin promoter (Gmubi), with an intron-GFP marker for detection of transgenic roots, may have widespread use in legume biology. Studies in which expression of the *Rhg1* locus LRR-kinase gene from different resistance sources was either reduced or complemented did not reveal significant impacts on SCN resistance.

Introduction

Soybean cyst nematode (SCN, *Heterodera glycines*) is an obligate, sedentary endoparasite that is consistently the most damaging pest of soybean in the U.S. (Wrather and Koenning, 2006). Once SCN is present in a field it cannot feasibly be eradicated.

The SCN life cycle consists of five stages. After the first molt within the egg, SCN second stage juveniles (J2) hatch, move through the soil, penetrate roots and move toward the vascular cylinder (Niblack et al., 2006; Ithal et al., 2007). Migratory juveniles select a host cell in the cortex, endodermis, or pericycle and induce host cell fusion as part of the formation of a permanent feeding site called a syncytium. At this point the nematode becomes sedentary and differentiates to the third (J3) and fourth (J4) juvenile stages and then matures to an adult female or male. Males undergo a metamorphosis to resume a vermiform shape at the J4 stage and migrate back out of the root to fertilize adult females. Following fertilization, the female produces eggs, most of which remain inside the body. After dying, the female body develops into a hardened cyst that encases the eggs. At 25^{o} C, some nematodes reach the adult stage 12 days after entering roots, and most become adults by 30 days post-infection (Lauritis et al., 1983).

Soybean cyst nematodes infect and grow in the roots of both resistant and susceptible cultivars (Niblack et al., 2006; Davis et al., 2008). Nematode growth and development depends on the successful establishment and maintenance of a syncytium, and impairment of the syncytium can give outcomes that range from reduced growth and reproduction to death. The available SCN resistance in soybean

is partial, and can be observed as a reduced number of females that develop compared to the number that develop on similarly inoculated susceptible cultivar controls (Halbrendt et al., 1992). In intact soybean plants, resistance is often expressed as the female index (FI): the number of fully developed females (cysts) on the tested soybean genotype divided by the number of females on a susceptible standard (Niblack et al., 2006; Colgrove and Niblack, 2008). Soybean cultivars are generally classified as strongly resistant to SCN if the FI is less than 10%; partial levels of resistance can also be useful (Schmitt and Shannon, 1992).

The SCN resistance in current commercially grown soybeans is derived from a very small number of sources (Brucker et al., 2005). These sources include 'Peking' (PI 548402), PI 88788 and PI 437654, which have each been shown to carry resistance loci effective against multiple nematode races (Concibido et al., 1997; Concibido et al., 2004). Inheritance of resistance to SCN was first reported in the 'Peking' plant introduction, and three genes for resistance (*rhg1*–*rhg3*) were assigned and initially classified as recessive (Caldwell et al., 1960). Of the resistance sources, PI 88788 has been the most widely used in breeding programs. More than 95% of the SCN resistant cultivars available for planting in Illinois during 2009 received their resistance from this PI (Shier, 2008).

The *Rhg1* locus has been shown to have the greatest impact on SCN development in several resistance sources including Peking, PI 88788, PI 437654, PI 209332 and PI 90763. This locus provides resistance to many common SCN populations such as Hg-type 0 (race 3) (Concibido et al., 2004). Multiple research

groups have mapped the *Rhg1* locus to a sub-telomeric region on chromosome 18 (Webb et al., 1995; Concibido et al., 1997; Cregan et al., 1999), to a location approximately 0.4 centimorgans (cM) from the simple sequence repeat (SSR) marker Satt309 (Cregan et al., 1999) (chromosome 18 was formerly known as linkage group G; http://www.phytozome.net/ (Schmutz et al., 2010)). Although originally reported as a recessive locus, "*rhg1"* has more recently been characterized as exhibiting incomplete dominance. Soybean lines heterozygous for resistant and susceptible alleles at the *Rhg1* locus often allow SCN cyst formation at a rate intermediate between that of plants genotyped with Satt309 as homozygous resistant or homozygous susceptible at the *Rhg1* locus (Lightfoot and Meksem, 2002; Ruben et al., 2006; Kim et al., 2010). A second QTL (*Rhg4*) has been identified as being necessary for full resistance to some SCN populations in Peking and in PI 437654, but not PI 88788 or PI 209332 (Brucker et al., 2005; Niblack et al., 2008). *Rhg4* exhibits dominant gene action, and in Peking-derived material the relevant alleles of both *Rhg1* and *Rhg4* are necessary to exhibit the full resistance phenotype. Other loci that make smaller and/or more race-specific contributions to SCN resistance have also been identified throughout the soybean genome, but often a given locus was identified in only one study (Riggs and Schmitt, 1988; Qiu et al., 1999; Concibido et al., 2004; Guo et al., 2006; Winter et al., 2007; Wu et al., 2009).

Cytological studies suggest that Peking-type resistance displays host cell necrosis and cell wall appositions not seen in PI 88788 type resistances in response to SCN Hg-type 0 (Mahalingam and Skorupska, 1996; Afzal et al., 2009). The Peking

and PI 88788 *Rhg1* sources also exhibit distinct differential behaviors in their strength of resistance against particular SCN test populations, suggesting at least partially different mechanisms in the SCN resistance controlled by different *Rhg1* alleles (Brucker et al., 2005; Klink et al., 2009).

Two groups first filed applications with the U.S. Patent Office in 2000 identifying apparent leucine-rich repeat transmembrane receptor-kinase (LRR-kinase) genes currently annotated as Glyma18g02680.1 and Glyma08g11350.1 (http://www.phytozome.net/) (Schmutz et al., 2010) as the likely SCN resistance genes at both *rhg1* and *Rhg4* (Hauge et al., 2001; Lightfoot and Meksem, 2002), see also (Hauge et al., 2006; Ruben et al., 2006). Sequences supporting the claims were released to Genbank between 2000 and 2005. The basis for these claims was the presence of these genes at the *Rhg1* and *Rhg4* loci, their similarity to the known rice bacterial blight resistance gene *Xa21*, and the presence of derived amino acid sequence differences between the alleles from resistant and susceptible plant genotypes. However, a decade later, no functional evidence for a role of these LRR-kinase genes in SCN resistance has been reported in a peer-reviewed forum. There are numerous reasons why identification of the *Rhg1* gene that confers SCN resistance is a high priority, including the extreme economic significance of SCN for yield loss, the major reliance on the *Rhg1* locus in commercial soybean breeding, the detection of SCN populations that overcome currently available *Rhg1*-mediated resistance, the potential to respond to these challenges with engineered

improvements in *Rhg1*, and basic scientific interest in the nature of plant resistance to SCN.

An experimentally tractable transgenic assay system is a crucial component for the identification, study and manipulation of genes controlling SCN resistance as well as many other soybean traits. Generation of transgenic fertile soybean lines is still a difficult and expensive process that requires close to a year to obtain transgenic seed lines. *Agrobacterium rhizogenes* has been used by many researchers as a transgene delivery system to study legume root biology, both in soybean and *Medicago truncatula* (Mazarei et al., 1996; Boisson-Dernier et al., 2001; Kereszt et al., 2007; Olhoft et al., 2007; Mrosk et al., 2009). Transgenic roots generated using *A. rhizogenes* retain the SCN resistance phenotypes of the parental soybean genotypes, and have been used to test genes that may impart resistance (Narayanan et al., 1999; Cho et al., 2000).

Function is often attributed to specific genes by the methods of gene mutation/positional cloning, phenotypic complementation via transformation with a cloned full-length gene, and/or by gene silencing. Experimental silencing of plant genes has generally been elicited using hairpin RNA-forming inverted repeat DNA constructs or by virus-induced gene silencing (Wesley et al., 2001; Baulcombe, 2004, 2005; Small, 2007). All silencing approaches are prone to incomplete penetrance (partial silencing, often for unpredictable reasons), but have nevertheless been useful for assigning function to specific genes, and for engineering useful traits. These gene silencing techniques have been used effectively in soybean and other

legumes (Ivashuta et al., 2005; Subramanian et al., 2005; Wasson et al., 2006; Graham et al., 2007; Nagamatsu et al., 2007; Constantin et al., 2008; Nagamatsu et al., 2009). With the more recent discovery of endogenous microRNAs as a major mode of gene regulation in many eukaryotes, artificial microRNA (amiRNA) methods have been developed for investigator-initiated silencing of target genes (Schwab et al., 2005; Schwab et al., 2006; Warthmann et al., 2008). Potential advantages of amiRNAs may include better penetrance, absence of undesired phenotypes associated with VIGS virus infections, and the capacity to limit off-target silencing of related sequences by elicitation with constructs specific for very short (19-24 bp) target sequences. However, use of amiRNA technology with soybean, *M. truncatula*, *Phaseolus* or other legumes has not been reported.

For the present study we refined assays that test SCN resistance in transgenic roots generated with *A. rhizogenes*. A nematode demographic assay was developed that discriminates resistant and susceptible responses by monitoring the infecting population for progression through nematode life stages. We also developed an amiRNA vector system for induction of gene silencing in legume roots using *A. rhizogenes* assay systems. We used these tools to investigate the impact on SCN resistance of the LRR-kinase gene from the *Rhg1* genomic region (the candidate SCN resistance gene). Our experiments expressing the LRR-kinase from a resistant (Peking/PI 437654-source) *Rhg1* locus in susceptible test lines, with or without the resistant allele at *Rhg4*, and silencing the LRR-kinase (PI 88788 source) in resistant lines, provided no evidence for a contribution of this gene to SCN resistance.

2.3. Results

2.3.1. Scoring resistance using a hairy root/nematode demographics assay

In soybean, *Medicago truncatula* and other plants for which stable transformation is difficult, transgenic "hairy roots" generated through *A. rhizogenes*-mediated transformation are often used to investigate gene function. A number of groups have previously shown that the resistance response to SCN in *A. rhizogenes*-derived roots correlates well with the SCN resistance phenotypes of intact soybean plants (Narayanan et al., 1999; Cho et al., 2000). Cyst formation is typically counted after 4 weeks but because the root tissues used often exhibit significantly reduced vigor 3-4 weeks into such assays, and to expedite and possibly improve resistance scoring, we investigated the scoring of resistance phenotypes at an earlier time point. Host resistance to SCN is often manifested as arrested development at specific life stages (Halbrendt et al., 1992; Narayanan et al., 1999), so we developed a "nematode demographics assay" in which SCN-infested transgenic root sections were stained with acid fuchsin approximately 2 weeks after inoculation with nematodes. Using a stereo dissecting microscope, the number of nematodes at each growth stage (J2, J3, or J4 + adult) was recorded for each root segment (Figure 1A, B, C). Data are expressed as a ratio: the number of nematodes within a root segment with an appearance resembling the J4 + adult, or J3 + J4 + adult, growth stage divided by the total number of nematodes (J2 + J3 + J4 + adult) within the root segment. Relative SCN resistance or susceptibility is determined by comparison to the demographic

data for known SCN-resistant or SCN-susceptible host genotypes tested within the same experiment. The results of two such assays are shown in Figure 1D and 1E. In these and other experiments (below and data not shown) there was a reproducible significant difference between the SCN resistance phenotypes determined for known SCN-susceptible and SCN-resistant control genotypes.

2.3.2. No elevation of SCN resistance observed in susceptible genotypes transformed with the *Rhg1* locus LRR-kinase gene from resistant germplasm.

The *Rhg1* locus of soybean has been shown in numerous studies to be the QTL with the greatest phenotypic impact on SCN resistance, and previous patents and other reports have indicated but not demonstrated that the Glyma18g02680.1 gene at the *Rhg1* locus that encodes an apparent leucine-rich repeat transmembrane receptor kinase (LRR-K) is the source of this SCN resistance. We isolated this LRR-K gene from the *Rhg1* locus using genomic DNA from the soybean cultivar Ina that we confirmed to exhibit SCN resistance in a standard whole plant/cyst formation assay immediately prior to its use in DNA isolation. Ina was used because of available progeny lines that are described below; Ina has Peking, PI 437654 and PI 88788 in its pedigree but carries an LRR-K at the *Rhg1* locus that is identical in sequence to the Peking allele and also similar to the PI437654 allele. Multiple independent PCR products were isolated and sequenced, and the sequences were compared to each other and to the Peking sequence published in (Hauge et al., 2001) to identify the

best clone. An *Agrobacterium*-compatible binary vector pGWB1::erGFP7INT carrying a screenable intron-GFP marker was constructed to transform soybean with candidate SCN resistance genes in full-gene complementation experiments. An insert of 6.9 kb was added to pGWB1::erGFP7INT and included approximately 2 kb of genomic upstream DNA, the predicted 2757 bp LRR-K coding region + intron, and approximately 800bp of downstream DNA. SCN-susceptible E98076 was then transformed with this construct to test the impact of this LRR-K allele on SCN development. Using the nematode demographics assay described above, roots from soybean line E98076 transformed with the *Rhg1* locus LRR-K did not display increased resistance to SCN; nematode development progressed similarly to what was observed on E98076 susceptible control roots transformed with the empty pGWB1::erGFP7INT vector (Figure 1D). E98076 transformed with either an empty plasmid vector or with the *Rhg1* locus LRR-K were, however, both significantly more susceptible to SCN than the resistant line Ina ($p<0.001$).

The *Rhg4* locus in Peking and PI 437654 also contributes to the full SCN resistance phenotype (Concibido et al., 2004). SCN-susceptible soybean genotypes with endogenous SCN-susceptible alleles at both *Rhg1* and *Rhg4*, which are then transformed with a resistant allele only of *Rhg1*, might not display a resistant phenotype. To further test the impact of the *Rhg1* locus LRR-K gene on SCN development, *rhg1*$^{-}$ lines carrying the SCN-resistant Peking *Rhg4* locus were tested by complementation. Microsatellite markers linked to *Rhg1* and *Rhg4* were used to identify Ina × E98076 progeny lines with a (-/-, +/+) genotype

(*rhg1*$_{E98076}$/*rhg1*$_{E98076}$, *Rhg4*$_{Ina}$/*Rhg4*$_{Ina}$). These lines were transformed with the cloned Ina *Rhg1*-locus LRR-K gene and assayed for SCN development. In the combined results from two independent experiments, no significant change in the resistance to SCN was detected after transformation of these genotypes with the Ina (Peking)-derived *Rhg1* locus LRR-K gene (Figure 1E). The two *rhg1*$^{-}$ lines transformed with an empty vector were significantly more susceptible to SCN than resistant Ina ($p<0.01$), while the *rhg1*$_{E98076}$/*rhg1*$_{E98076}$, *Rhg4*$_{Ina}$/*Rhg4*$_{Ina}$ genotype transformed with the *Rhg1* locus LRR-K gene was not significantly more resistant than the same genotype transformed with empty vector. The mean resistance of roots transformed with the Ina LRR-K gene may have been slightly greater but not to a significant extent, in contrast to the statistically significant difference between the Ina and E98076 control lines. When the nematode demographics data for these same experiments were expressed as [(J3+J4)/total], the results for *rhg1*$_{E98076}$/*rhg1*$_{E98076}$, *Rhg4*$_{Ina}$/*Rhg4*$_{Ina}$ genotypes transformed with empty vector or vector + *Rhg1* locus LRR-K gene were even more similar (mean ± std error 0.59 ± 0.09 for empty vector and 0.49 ± 0.09 for LRR-K Ina; P = 0.85 for ANOVA Tukey simultaneous tests).

2.3.3. Construction of an amiRNA gene silencing vector

We sought to use gene silencing as an independent means to test for contributions of the *Rhg1* locus LRR-kinase gene toward SCN resistance. To facilitate this work, we

constructed a new vector for gene silencing in transgenic soybean roots. The binary vector pSM103 (pCAMBIA1300::35SP-erGFP7INT-NOS, Gmubi promoter, mi319RNAa) was designed to meet the following requirements: 1) efficient cloning of artificial microRNAs, 2) constitutive strong amiRNA expression in legume roots as well as other tissues (Hernandez-Garcia et al., 2009), 3) reliable identification of transgenic roots by GFP screening using an intron-containing GFP that is not expressed in bacteria, 4) competence for transferring inserted DNA into plant genomes via *Agrobacterium*, and 5) high transformation efficiency. The steps taken to develop the vector are shown in Figure 2, as the intermediate constructs may also be of use and are available upon request. To initially confirm GFP7 expression *in planta,* pSM101 and pSM103 were infiltrated into *N. bentamiana* leaves using *A. tumefaciens GV3101*. Positive and negative controls (pGWB1::35SP-erGFP7INT-NOS, and *A. tumefaciens GV3101* without any vector) did or did not produce green fluorescence respectively, as expected (Figure 3A and 3B). erGFP7INT expression was phenotypically detected with use of vectors pSM101 and pSM103 (Figure 3C and 3D).

2.3.4. Silencing of the *Rhg1* locus LRR-Kinase in transgenic roots.

For silencing experiments, the *Rhg1* locus from the more widely used PI 88788 source of *Rhg1* was tested using the soybean cultivar Fayette. Cotyledons of Fayette were transformed with *A. rhizogenes* containing pSM103 derivatives carrying one of

two different amiRNAs that each were designed to specifically target the Gm18 *Rhg1* locus LRR-K gene. To determine the effectiveness of the two amiRNAs, RNA was extracted from three independent transgenic hairy roots for each construct. Semi-quantitative RT-PCR was performed in each of two independent biological replicate experiments as shown in Figure 4A. The semi-quantitative RT-PCR was also performed using Fayette GFP-negative hairy roots to determine the basal expression level of the *Rhg1* locus LRR-kinase gene, and using Fayette GFP-positive roots to verify that unmodified pSM103 does not interfere with expression of the *Rhg1* locus LRR-kinase gene. No significant differences were found between the two Fayette controls (Figure 4A and 4B). ANOVA was utilized to evaluate differences in expression of the *Rhg1* locus LRR-K gene. As shown in Figure 4A and B, at 35 cycles only the "amiRNA LRR-K I" construct was able to substantially reduce the expression level (mRNA abundance) of the *Rhg1* locus LRR-kinase gene in transgenic roots (ANOVA $p<0.05$). No statistical difference was found for the amiRNA LRR-K II, but a trend of partial down regulation was observed (Figure 4B). The actin control did not show a significant difference among the samples.

2.3.5. Impacts on Transformation Efficiency and Root Development

Transformation efficiency was evaluated 3 weeks after *A. rhizogenes* application, with GFP used as a visual marker for successful transformation. Transformation frequency was calculated as the number of cotyledons with at least

one transgenic hairy root divided by the total number of inoculated cotyledons. Fayette and Williams 82 transformed with pSM101, pSM103 and pGWB1::35SP-erGFP7INT showed transformation efficiencies ranging from 43.33% to 50%. No significant difference in transformation efficiency was observed between the two genotypes transformed with the different control vectors (Figure 5B). However, when Fayette cotyledons were transformed with amiRNA GmLRR-K I and amiRNA GmLRR-K II, decreased transformation efficiency was observed. Cotyledons transformed with the two LRR-kinase amiRNA constructs developed hairy roots roughly two weeks later, and had lower transformation efficiencies than the controls. In particular, the amiRNA GmLRR-K I gave the lowest transformation efficiency (8.37%) and the resulting roots grew more slowly (Figure 5A and B). Based on two independent transformation experiments using 20 cotyledons of each genotype, a Chi-square test revealed no significant different between Williams and Fayette controls for root transformation efficiency (χ^2 P = 0.518), while a significant difference was observed for amiRNAGmLRR-K I (χ^2 P = 3.56e-10) and amiRNA GmLRR-K II (χ^2 P = 1.156e-4) relative to the expected rate established from Fayette controls.

To evaluate potential artifactual impacts of the amiRNA vector in plants lacking the target gene, a transformation experiment was performed using *Medicago truncatula.* The Fayette TC204550 sequence has only 81% identity with *M. truncatula* Chromosome 5 clone mth2-155g22 (CU302335). *M. truncatula* seeds were transformed with *A. rhizogenes* to deliver the same silencing constructs tested

in soybean. Three plates, each with 12 germinated plants, were used for each construct. Three weeks after the treatment, no difference in root development was observed between *M. trucatula* seedlings transformed with the two amiRNA GmLRR-K constructs in comparison to the unmodified pSM103 vector. The transformation efficiencies were not substantially different, ranging from 10% to 30% amongst the plates (data not shown). These results indicate the amiRNAs designed to silence the soybean *Rhg1* locus LRR-kinase gene have a root development impact specifically in soybeans.

2.3.6. No Significant Impact on SCN Resistance after Partial Silencing of the *Rhg1* locus LRR-kinase

Healthy transgenic GFP-positive roots expressing amiRNA constructs that target the *Rhg1* locus LRR-K gene (or empty vector controls) were sub-cultured by propagating the young growing tips (2-3 cm), and were infected with ~250 sterile J2 SCN for each root. The nematode demographics assay was then used to detect relative SCN resistance. The experiment was repeated with three independent biological replicates each using 4-10 separate transgenic hairy roots of each treatment-construct. Two weeks after nematode inoculation, all nematode stages (except eggs) were observed in the transgenic roots tested, and the number of nematodes at each stage was recorded. Roots carrying five or fewer nematodes were excluded from further analysis. As expected, Fayette control roots allowed

significantly fewer nematodes to molt and progress to an apparent J3 or J4/adult female stage within 14 days relative to Williams 82 control roots (Figure 6; $p = 0.007$). In the key comparison to test if silencing of the *Rhg1* locus LRR-kinase gene altered resistance to SCN, no statistical difference in nematode development was observed between Fayette and Fayette carrying amiRNA GmLRR-K I or amiRNA Gm-K II (Figure 6).

2.4. Discussion

As one part of this study, we report use of amiRNA technology to knock down expression of an endogenous gene in soybean. Like RNAi (hairpin RNA or VIGS-mediated silencing), miRNAs and amiRNAs in plants reduce gene expression primarily by eliciting cleavage of homologous RNA transcripts. The primary benefit of adopting amiRNA technology is specificity (Schwab et al., 2005; Schwab et al., 2006; Warthmann et al., 2008). Longer hairpin RNA gene silencing constructs (often 100-800 bp) are cleaved into multiple subfragments that, with the known tolerance for mismatches within the RNA-induced silencing complex, can cause unintended silencing of non-target genes. This can readily remain undetected by the investigator. The target recognition portion of miRNAs and amiRNAs are very short and are strand specific, and plant miRNA-mediated silencing is more sequence-specific than animal miRNA-mediated silencing, all of which greatly reduces the potential for unintended off-target effects on other genes. Based on known target selectivity behaviors (see for example http://wmd3.weigelworld.org/ and (Schwab et al., 2005), the amiRNA can be designed to be gene specific - even targeting specific splice forms - but like other gene silencing methods, amiRNAs can alternatively be designed to silence multiple similar genes if an appropriate shared target sequence is used.

We constructed the vector pSM103 to address three goals: ready acceptance of newly generated amiRNA constructs into an *Agrobacterium* plant transformation vector, strong expression of the amiRNA in multiple tissues in soybean, and presence of an intron-disrupted GFP to allow identification of transgenic tissues. Each of these ideas has ample precedent but they had not been brought together for soybean. Gmubi is a strongly expressed promoter - offering potential improvements over CaMV 35S and other promoters for use in soybean and other legumes (Chiera et al., 2007; Hernandez-Garcia et al., 2009). Although the Arabidopsis miRNA319a has been an effective template for successful amiRNAs used in tomato, tobacco, *Physcomitrella patens* and other species (Alvarez et al., 2006; Khraiwesh et al., 2008), and was effective in the present use in soybean, it is possible in the future that improved silencing with amiRNA in soybean may be achievable by using one of the many endogenous soybean miRNAs as the template (Zhang et al., 2008; Zhang et al., 2010). Another permutation for future investigation may be the use of highly specific promoters to drive expression of the amiRNA, to elicit silencing in a more limited set of cells, tissues or developmental stages. However, the success of amiRNA was demonstrated in the present study by the successful reduction in target mRNA abundance and by the elicitation of root development impacts in lines carrying amiRNA LRR-K I. In current work with other soybean genes, we are finding that roughly half of the amiRNAs that we construct cause detectable reduction in target mRNA abundance (unpublished), as is commonly observed in most

laboratories that work with hairpin RNAs, amiRNAs and other engineered silencing methods.

The nematode demographics assay was developed as a modification of existing protocols that monitor SCN development on transgenic soybean roots generated using *A. rhizogenes* (Narayanan et al., 1999; Cho et al., 2000). There were multiple motivations for developing the demographics approach. First and most important, we hypothesized that a census of the developmental stages achieved by the infecting SCN population mid-way through the infection life-cycle might be more sensitive than an end-stage cyst count in detecting differences in the expression SCN resistance, which is a quantitative trait. Previous reports indicate that morphological differences between nematode development on SCN-resistant and SCN-susceptible soybeans are evident relatively early - as soon as 5 days after infection (Endo, 1965; Halbrendt et al., 1992; Narayanan et al., 1999; Colgrove and Niblack, 2008). Gene expression profiling work has even suggested host gene expression differences between resistant and susceptible interactions in the first 12 hours (Klink et al., 2007). A second motivation for the demographics assay was that root segments often became less healthy in appearance after 3-4 weeks in culture. Root tips remain vigorous and continue to grow, but under our conditions, older segments such as the segments present at the time of nematode infestation start to exhibit notable decline approximately three weeks after exposure to nematodes. Data collection at an earlier time point reduces this concern. A third useful attribute of the demographics assay is that results become available two weeks earlier,

compared to cyst counts at 30 days post-infection. Lastly, when first establishing the assay in our laboratory, slow-growing fungal contaminants sometimes emerged in the third or fourth week (usually introduced by application of clean but not fully sterile nematodes). Although this issue was addressed primarily by more effective nematode surface-sterilization or use of axenic nematode cultures, use of an assay that terminates two weeks after SCN application reduced contamination issues (note that more thorough surface sterilization is lethal to a high percentage of the nematodes and introduces its own potential problems). The validity of the demographics scoring approach was supported by the significant differences in SCN development we observed between known resistant and susceptible plant genotypes. Acid fuchsin staining and nematode counting do require some additional effort, however, and previously published protocols that count cyst formation after 28-35 days remain as an acceptable alternative.

In addition to reporting amiRNA use for gene silencing in soybean, a vector to facilitate this approach, and the nematode demographics assay for tests of genes that may contribute to soybean SCN resistance, we also report negative data regarding impacts of the *Rhg1*-locus LRR-K gene on SCN resistance. Using complementation with a full-length gene under control of its native promoter, and gene knockdown using amiRNA technology, we were unable to detect a contribution to SCN resistance by the *Rhg1*-locus LRR-K gene from either Peking/PI 437654 or PI 88788, the most commonly used sources of SCN resistance in U.S. soybean cultivars. An impact on root development was observed when amiRNA for the LRR-K gene

were expressed in transgenic roots, and effects on root development have also been preliminarily reported by the Lightfoot research group (http://hdl.handle.net/10101/npre.2008.2726.1). It remains possible that this *Rhg1*-locus LRR-K gene may in some way impact SCN infection of or development on soybean roots, but no such effect was detected in our studies.

The *Rhg1* locus is of extremely high economic significance to soybean cultivation in the United States, and has received attention from multiple research groups. It is notable that more than ten years after the filing of patent applications claiming SCN resistance function for the *Rhg1*-locus LRR-K gene, no groups have published evidence (beyond its presence as one of many candidate genes at the *Rhg1* locus) for a functional contribution to SCN resistance by the *Rhg1*-locus LRR-K gene. We note that, simultaneous with the present study, a group led by Drs. Khalid Meksem and Melissa Mitchum has completed tests for functional contributions of the *Rhg4*-locus LRR-K gene to SCN resistance and, similar to our work, has not detected a contribution for this *Rhg4* candidate gene (K. Meksem, personal communication). We are also aware of work led by Dr. Brian Diers (Kim et al., submitted), in which fine-structure genetic mapping of the *Rhg1* locus has revealed chromosomal recombination events that genetically separate *rhg1-b* (PI 88788-source) SCN resistance function from the *Rhg1*-locus LRR-K gene. Hence evidence is accumulating that the *Rhg1* locus LRR-K genes seem unlikely to be the primary source of SCN resistance polymorphism. A tone of caution is appropriate, however, as our present study only provides negative data on this subject. For example, gene

silencing is often partial (whether mediated by VIGS, hairpin RNA-generating constucts or amiRNA), and the expression knockdown we achieved with amiRNA may have been insufficient to detectably block hypothesized SCN resistance contributions (although the other phenotypic impacts we observed with the amiRNA LRR-K I suggest that the gene expression knockdown achieved was functionally significant). As another example, the quantitative nature of the functional contribution of any given gene to SCN resistance may be difficult to detect in *A. rhizogenes*-generated soybean roots. Hence we are cautious with our conclusion, but we can state that when SCN-susceptible sources were complemented with an *Rhg1*-locus LRR-K allele from an SCN-resistant source, or when expression of the gene in resistant plants was reduced by amiRNA-mediated silencing, no significant impacts on SCN resistance were detected.

2.5. Conclusions

Our findings suggest that the nematode demographics assay can expedite testing of transgenic roots for SCN resistance. amiRNAs may have widespread use in legume biology, and we have developed the vector pSM103 that drives interchangeable amiRNA constructs through a Gmubi promoter, with an intron-GFP marker for detection of transgenic roots. Studies in which expression of the *Rhg1* locus LRR-kinase gene from different resistance sources was either reduced or complemented did not reveal significant impacts on SCN resistance, suggesting that one or more

other genes at the *Rhg1* locus may be the primary determinant of differences in SCN resistance among different soybean types.

2.6. Methods

2.6.1. Agrobacterium cultures

A. tumefaciens GV3101 and *A. rhizogenes* strain Arqua1 were transformed by freeze-thaw (Hofgen and Willmitzer, 1988) or by electroporation (Wise et al., 2006). The cells were plated on selective media with the appropriate antibiotic and incubated at 28°C for about 2 days. *A. rhizogenes* strain Arqua1 was received from Dr. Jean-Michel Ane, University of Wisconsin Madison.

2.6.2. Soybean lines

Soybean cultivars Fayette and Ina (SCN-resistant lines), and Williams 82 and E98076 (susceptible lines) were used; E98076 is an SCN-susceptible experimental line from Dechun Wang, Michigan State University. Genotyped Ina x E98076 progeny lines were kindly provided by Dr. Brian Diers, University of Illinois at Urbana-Champaign. Those lines were inferred to be homozygous for $Rhg1_{\text{Ina}}$ or $rhg1_{\text{E98076}}$, and for $Rhg4_{\text{Ina,}}$ or $rhg4_{\text{E98076}}$, using micro satellite markers Satt309 and Satt424 (Song et al., 2004); http://soybeanbreederstoolbox.org/. DNA sequencing was used to confirm nucleotide identity between the *Rhg1*-locus LRR-kinase gene from cultivars Ina and Peking.

2.6.3. Surface sterilization

After visual screening for obvious signs of fungal or viral contamination, soybean seeds were surface-sterilized for 16-20 h in a desiccator jar with chlorine gas by adding, to a beaker within the jar, 3.5 ml 12N HCl into 100 ml household bleach (6% sodium hypochlorite) immediately before closing the jar. At least 20 seeds per experiment were plated onto germination media (Gamborg's B5 salts (3.1g/L), 2% sucrose, 1X Gamborg's B5 vitamins, 7% Noble agar, pH 5.8) in 100x25 mm Petri plates. Plates were wrapped with Micropore tape (3M, St. Paul, MN) and incubated at 26^{o}C in a growth chamber (18/6 light/dark hours) for about a week. *Medicago truncatula* seeds (approximately 150 per batch) were scarified in concentrated sulfuric acid for 8 min, washed 3 times in sterile water, plated on 1% deionized water agar plates, and then stratified for 24h or 48 h at 4°C and germinated by incubating at room temperature overnight (Peiter et al., 2007).

2.6.4. Soybean and *Medicago* transformation

Soybean cotyledons were harvested about 7 days after germination by gently twisting them off the hypocotyls. With a sterile forceps and scalpel, several shallow slices were made across the abaxial surface of the cotyledons after dipping the scalpel in *A. rhizogenes* suspension (OD600 0.7 in Luria broth). The cotyledons were then placed abaxial-side down on a co-culture medium (0.31g/L Gamborg's B5 salts,

3% sucrose, 1X Gamborg's B5 vitamins, 0.4g/L L-cysteine, 0.154g/L dithiothreitol, 0.245g/L sodium thiosulfate, 40mg/L acetosyringone, 5% Noble agar, pH 5.4) in 100x15mm Petri plates with a piece of 70mm filter paper (Whatman, Piscataway, NJ) on the surface of the agar to prevent *A. rhizogenes* from overgrowing. Plates were wrapped with parafilm and incubated in the dark at room temperature for three days. The explants were then transferred to a hairy root medium (HRM) of 4.3g/L MS salts (Sigma Co., St. Louis, MO), 3% sucrose, 1X Gamborg's B5 vitamins (Sigma Co. St. Louis, MO), 7% Noble agar, 0.15g/L cefotaxime, 0.15g/L carbenicillin, pH 5.6 in 100x15mm Petri plates, wounded side up. Plates were wrapped with Micropore tape and incubated in the dark at room temperature until roots emerged, usually in around 3 weeks.

To generate *M. truncatula* with transgenic roots, *A. rhizogenes* Arqua1 carrying pSM103 with one of the two different amiRNA for the *Rhg1* GmLRR-Kinase (amiRNA GmLRR-KI and GmLRR-KII), or the empty vector, were used for *M. truncatula* transformation (Boisson-Dernier et al., 2001). Following transformation, seedlings were grown for 3 weeks. The bottom part of the plates (corresponding to the roots) was covered with aluminum foil. The plates and the *M. trucatula* composite plants were watered with sterile water every other day.

2.6.5. Transgenic root identification and propagation

Soybean or *M. truncatula* transgenic roots were detected based on GFP expression, using a fluorescence stereomicroscope (LEICA MZ FL III with GFP2 filter). Transgenic soybean root tips were cut into 2-3 cm segments and transferred to HRM. Roots that were expressing incomplete strips of fluorescence (chimeras) or exhibiting overall low levels of GFP fluorescence were avoided. GFP-positive as well as negative control roots were generally used within two months of the initial transformation event, after maintenance by transferring 2-3 cm root tips every two weeks. Independent transgenic events, generated from different inoculation sites or different cotyledons, were maintained separately for use in RT-PCR and nematode demographic assays.

2.6.6. Cloning of candidate *Rhg1* LRR-kinase gene

The LRR-kinase gene Glyma18g02680.1 from the *Rhg1* locus was cloned from Ina genomic DNA. Primers were designed based on BAC sequences from soybean cultivar A3244 (Hauge et al., 2001), using the Primer3 tool. The Roche Expand Long™ (Roche Applied Science, Indianapolis, IN) kit and protocol were used with primers

Rhg1LRR-KF: TTCACCCGTGATACATGTTAATTC;

Rhg1LRR-KR: GGAACTTGGAAGTCATTATCTTGG

to amplify a 6.9 kb PCR product from the *Rhg1* locus that includes >2 kb of upstream sequence, the coding region, and 0.8 kb of downstream sequence, which was then cloned into the Invitrogen plasmid pCR8/GW/TOPO according to the manufacturer's instructions. Multiple independent clones were isolated and sequenced in order to avoid clones with PCR-induced mutations. The binary vector pGWB1::erGFP7INT for full-length gene complementation experiments was constructed from pGWB1 (Nakagawa et al., 2007) by addition of the 35S-promoter/erGFP7INT/NOS-terminator of pLMNC95 (Mankin and Thompson, 2001), taken from a pCAMBIA 1300 series vector by digestion with EcoRI and HindIII and inserted into the HindIII site of pGWB1 after blunting. The LRR-kinase genes (in pCR8/GW/TOPO) from the *Rhg1* locus were moved in to pGWB1::erGFP7INT using the Invitrogen Gateway™ system (Invitrogen, Carlsbad, CA) after linearizing the pCR8/GW/TOPO-derived plasmids with XbaI. To obtain a negative control transformation vector lacking the Gateway *ccdB* negative selectable marker, a 167 bp segment of nonsense DNA was cloned and transferred into pGWB1::erGFP7INT.

2.6.7. Nematode maintenance

A Wisconsin SCN population (Hg-type 7), collected by A.E.M., was maintained on the susceptible soybean cultivar Williams 82. Seeds were germinated in a wet 70mm filter paper (Whatman) in a Petri plates for 4-5 days, planted in autoclaved 2:1 sand:soil mixture, inoculated with 2000-3000 eggs of *H. glycines* per container, and

grown in a water bath at 28°C in a 28°C greenhouse. Cysts were collected 6 weeks after infection, and were extracted from infested pots using sieves and centrifugation. Briefly, soil from infected pots was placed in a pitcher of water and the roots were massaged to dislodge attached cyst. The soil-cyst-water slurry was passed over a 595 µm - 250 µm sieve tower, and the mixture from the 250 µm sieve was backwashed into a 50mL plastic conical tube. The tubes were centrifuged at 2000 RPM for 4 minutes then the supernatant was poured off. A 60% sucrose solution was added to the tubes, stirred, and centrifuged at 2000 rpm for 2 min. Cysts in the supernatant were then collected over a 250 µm sieve. Collected cysts were stored at 4°C in plastic bags in damp sand.

2.6.8. Nematode demographics assay

For nematode demographics assays, *H. glycines* eggs were collected by breaking open cysts and collecting the eggs on a sieve stack consisting of 250 µm - 74 µm - 25 µm sieves (USA Standard Testing Sieve). Eggs were collected from the 25 µm sieve and rinsed. Eggs were placed in a hatch chamber (Schroeder and MacGuidwin, 2006) with 3 mM $ZnCl_2$ for hatching at room temperature in the dark for 5-6 days. Hatched J2 nematodes were surface-sterilized for 3 min in 0.001% mercuric chloride and washed three times with sterile distilled water (Sindhu et al., 2009), then suspended in 0.5% low-melting point agarose to facilitate even distribution. The number of active nematodes was counted using a hemacytometer at least one-

half hour after surface-sterilization and washing, and 250-500 active J2s were inoculated onto each fresh root segments, depending on the experiment (similar numbers of active J2 were applied to all roots within an experiment). Roots (2-3 cm vigorous new root segments closest to and including root tip) with nematodes were maintained on HRM media at 28^{o}C (in initial experiments including some experiments reported in Figure 1, roots were placed on cellophane on HRM media for 24 hr. starting just before nematodes were added, and then transferred to HRM without cellophane and incubated vertically for 48 hr., then incubated horizontally until experiment was complete). Nematode infection and development within root segments was monitored by clearing and staining with acid fuchsin as in (Byrd et al., 1983), typically 15 days post inoculation (dpi). The nematode demographic assay was then completed by recording the number of nematodes in each root with a morphology resembling either J2 (thin), J3 (sausage-shaped) or J4/adult female nematodes, as noted in text and figures.

2.6.9. Construction of amiRNA vector

To construct the pSM103 vector for amiRNA-mediated gene silencing in legume roots (Figure 2), the 35S-promoter/erGFP7INT/NOS-terminator construct (intron-GFP) from pLMNC95 (Mankin and Thompson, 2001) was mutagenized to remove XbaI and BamHI sites, and then cloned from pLMNC95 (ΔXbaI and ΔBamHI) into pCAMBIA 1300 after PCR-mediated addition of restrictions sites using the primers:
Plmnc95_Kpn1F: CCAGGTACCCAGGTCCCCAGATTAGCC;

Plmnc95EcoR1_R: GCCAGTGAATTCCCGATCTA.

The amiRNA constructs from pRS300 derivatives (see below) were added to pCAMBIA 1300::35SP-erGFP7INT-NOS using BamHI/PstI sites. The soybean polyubiquitin (gb|EU310508.1|) promoter (GmUbi) (Chiera et al., 2007; Hernandez-Garcia et al., 2009) was amplified from Fayette genomic DNA with added HindIII and PstI sites using the primers:

Ubi1_HindIII_Fa: CCAAAGCTTGGGCCCAATATAACAACGAC;

Ubi_PstI_Ra: CCACTGCAGCTGTCGAGTCAACAATCA.

The Gmubi product was cloned into pCR2.1 TOPO and then in pCAMBIA 1300::35SP-erGFP7INT-NOS::miR319a using HindIII/PstI sites.

2.6.10. Transient Expression Assay

The plasmids pGWB1::erGFP7INT, pSM101 and pSM103 were initially tested for GFP function by agroinfiltration of *A. tumefaciens* GV3101 into *Nicotiana benthamiana* leaves, along with GV3101 without binary plasmid as a negative control. *A. tumefaciens GV3101* strains carrying pGWB1::erGFP7INT, pSM101 and pSM103 were grown overnight at 28°C in 5ml LB medium with 2.5 µg/ml of rifampicin and 50 µg/ml of kanamycin (the negative control *A. tumefaciens* GV3101 without binary vector was grown only with rifampicin). *A. tumefaciens* cells were harvested by centrifugation and resuspended to OD_{600} 0.4 in MMA induction buffer (5 g MS salts, 20 g sucrose, 1.95 g MES per liter, 200 µM acetosyringone, pH to 5.6

with 1M NaOH. Bacterial suspensions were then incubated at room temperature for 1-3 hr prior to infiltration through the abaxial surface into the leaf, using a syringe with no needle. GFP fluorescence was investigated 3 days after infiltration in excised leaf sections that included the infiltrated zone and 1-3 cm of surrounding non-infiltrated tissue, using an Olympus BX-60 upright Fluorescence Microscope and MF filter.

2.6.11. Sequencing and amiRNA design.

EST TC204550 UP|Q8L3Y5 (Q8L3Y5) (Glycine max EST GmGI-12.0; wmd2 database http://wmd2.weigelworld.org/cgi-bin/mirnatools.pl) corresponding to the Glyma18g02680.1 putative LRR-kinase gene from the *Rhg1* locus was used to design PCR primers for isolation of the *Rhg1* locus receptor-like kinase-encoding allele from the resistant source Fayette, which carries the *Rhg1* resistance allele *rhg1-b* from PI 88788. The primers RTPCR_TC204550F CAA CTT CAA GGC CTC AGG AA; RTPCR_TC204550R GCT ACC CAA AGA AGC AGG AA, amplified a PCR product of 450bp. The DNA sequences of PCR products were determined by the dideoxy chain-termination method using ABI big dye cycle sequencing kit (3.1) and the ABI 377 automated sequencing service at the University of Wisconsin - Madison Biotechnology Center. Blast analysis of the Fayette version of TC204550 (the EST for this gene in earlier WMD versions) was conducted in March 2008 using the Phytozome database (http://www.phytozome.net/soybean). For amiRNA design,

the sequence of this Fayette LRR-kinase gene was submitted to http://wmd2.weigelworld.org. The wmd2 tool identified a total of 85 candidate amiRNA sequences to potentially silence the *Rhg1* locus LRR-kinase gene, with the first 31 candidates classified by wmd2 as having excellent (most promising) properties (data not shown). The two amiRNA target sequences selected for primer design to modify pRS300 for silencing of the *Rhg1* locus LRR-kinase gene were identical to Fayette chromosome 18 sequence as well as the paralogous chromosome 11 sequence from the Williams 82 genome: amiRNA GmLRR-KI: TAAGACTATAAGGGATTGCTG; and amiRNA LRR-KII: TAAGACTATAAGGGATTGCTC. The Arabidopsis miR319a sequence from pRS300 was used as a template to create three PCR products (a, b, c) with soybean target transcript specificity. These products were used as template for a fourth PCR using overlapping primers to create gene specific amiRNA (Schwab et al., 2006). The final PCR product (containing target transcript sequence) was cloned into pSM103 using BamHI-PstI restriction sites, replacing the miR319a sequence. Presence of the desired amiRNA was confirmed by DNA sequencing.

2.6.12. DNA and RNA extraction and RT-PCR

Soybean genomic DNA was extracted from young leaves using a CTAB method previously described (Murray and Thompson, 1980). Total RNA was extracted from independent transgenic hairy roots (about 3 cm long) that had been frozen in liquid nitrogen and stored at -80°C until use. Total RNA was isolated using Qiagen RNeasy

Mini Kit according to the manufacturer's instructions, and treated with RNase-free DNase I (Qiagen, Valencia, CA, U.S.A.). First-strand cDNA was synthesized from 250 ng of DNase-treated RNA using SuperScript III Reverse Transcriptase. The RNA concentration was determined using the NanoDrop-1000 Spectophotometer. Semi-quantitative RT-PCR was performed using 25 ng of cDNA. Primers designed from the Fayette *Rhg1* locus LRR-Kinase target, RTPCR_TC204550F CAACTTCAAGGCCTCAGGAA (forward) and RTPCR_TC204550R GCTACCCAAAGAAGCAGGAA (reverse), were used to amplify a 450bp product. As a control for equivalent starting pools of cDNA, the primers Actin forward GTTCTCTCCTTGTATGCAAGTG and Actin reverse CCAGACTCATCATATTCACCTTTAG were used to amplify a 700bp segment of soybean actin gene cDNA (emb|V00450.1|) The RT-PCR was performed in triplicate at 55°C, testing three different numbers of cycles (30-35-40) to confirm non-saturation of amplification. PCR products were visualized by ethidium bromide staining after separation on 1.5% agarose gels and band intensity was quantified using the Kodak 1D Image Analysis Software v3.6.

2.6.13. Statistical analysis

Data were analyzed by ANOVA using Minitab (v.14) with the General Linear Model and Tukey Simultaneous Test; nematode demographics data for the amiRNA experiments were analyzed using a mixed model (Mixed Procedure; SAS, vers. 9.1).

2.7. Author's Contributions

SM, ALH and DC planned, conducted and analyzed most of the experiments and were centrally involved in writing the manuscript. BWD developed soybean germplasm and provided significant ideas and critical review of the manuscript. AEM contributed nematode populations, methodological suggestions and biological insights. AFB conceived the overall project, analyzed results and planned experiments, and was a primary author of the manuscript. All authors read and approved the final manuscript.

2.8. Acknowledgments

This project was supported by grants to A.B. from the Wisconsin Agricultural Research Station, the Wisconsin Soybean Marketing Board and the United Soybean Board. We thank Jean-Michel Ane and Muthusubramanian Venkatashwaran for the Arqua1 strain, helpful discussions and assistance with Medicago experiments, Michelle Graham for early help in identifying regions of duplication between the Gm18 *Rhg1* locus and Gm11, and Melissa Mitchum for advice about SCN testing with *A. rhizogenes*-generated roots and for critical reading of the manuscript.

2.9. Figures

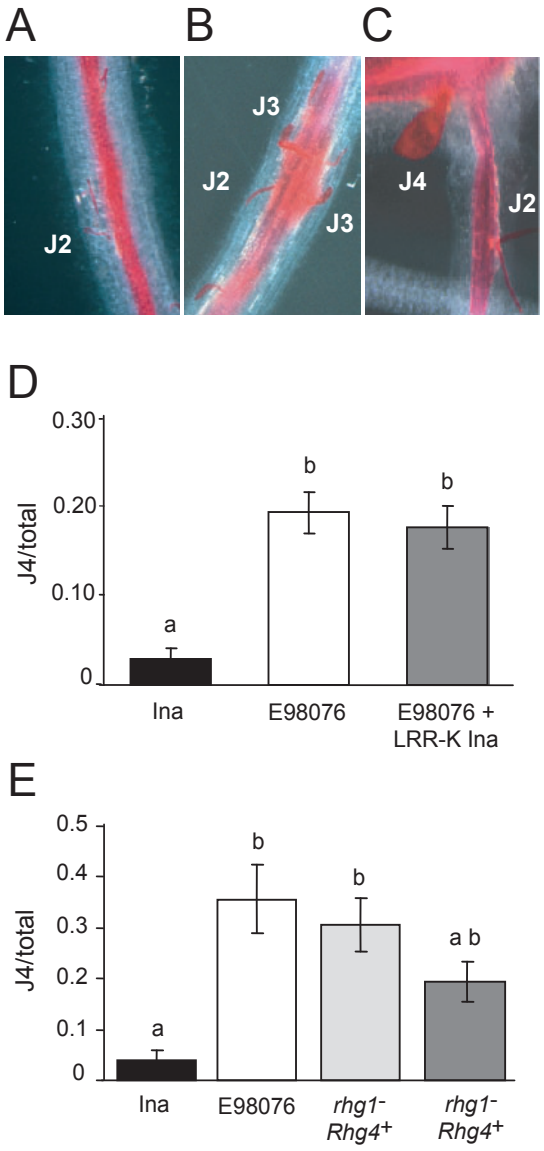

A
B
C
J2
J3
J2
J3
J4
J2
D
J4/total
0.30
0.20
0.10
0
a
b
b
Ina
E98076
E98076 +
LRR-K Ina
E
J4/total
0.5
0.4
0.3
0.2
0.1
0
a
b
b
a b
Ina
E98076
rhg1-
Rhg4+
rhg1-
Rhg4+

Figure 1.

Nematode demographics assay for SCN resistance in roots transformed with a full-length *Rhg1*-locus LRR-Kinase gene.

A, B, C, Representative examples of SCN developmental stages on transgenic soybean hairy roots 15 days after infection; cleared root sections were stained with acid fuchsin, which stains nematodes as well as root vascular bundles. **A,** Early

Figure 1 cont'd

parasitic second-stage juveniles (J2). **B,** Late parasitic J2s and early J3s. **C,** J4 or adult female. **D,** Ratio of total infecting nematode population that developed to J4 or adult female stage two weeks after inoculation onto *A. rhizogenens*-transformed roots of Ina, E98076 (both transformed with empty vector), or E98076 + LRR-K Ina (transformed with the Glyma18g02680.1 *Rhg1*-locus LRR-Kinase promoter/gene/terminator). Mean ± std error of mean; data combined from two independent experiments; bars with the same letter above are not significantly different (ANOVA, Tukey pairwise comparison, $p<0.001$). **E,** Similar to D, except that "*rhg1*$^{-}$ *Rhg4*$^{+}$ are putative *rhg1*$^{-}$/*rhg1*$^{-}$**;** *Rhg4*$^{+}$/*Rhg4*$^{+}$ progeny lines from Ina x E98076 that were genotyped as carrying the E98076 (susceptible) allele of a marker linked to *Rhg1* and the Ina (resistant) allele of a marker linked to *Rhg4*. All the lines were transformed with empty vector except "*rhg1*$^{-}$ *Rhg4*$^{+}$ + LRR-K Ina" was transformed with *Rhg1*-locus LRR-kinase gene (including native promoter and terminator). Mean ± std error of mean are from two independent experiments; bars with the same letter above are not significantly different (ANOVA, Tukey pairwise comparison, $p < 0.05$).

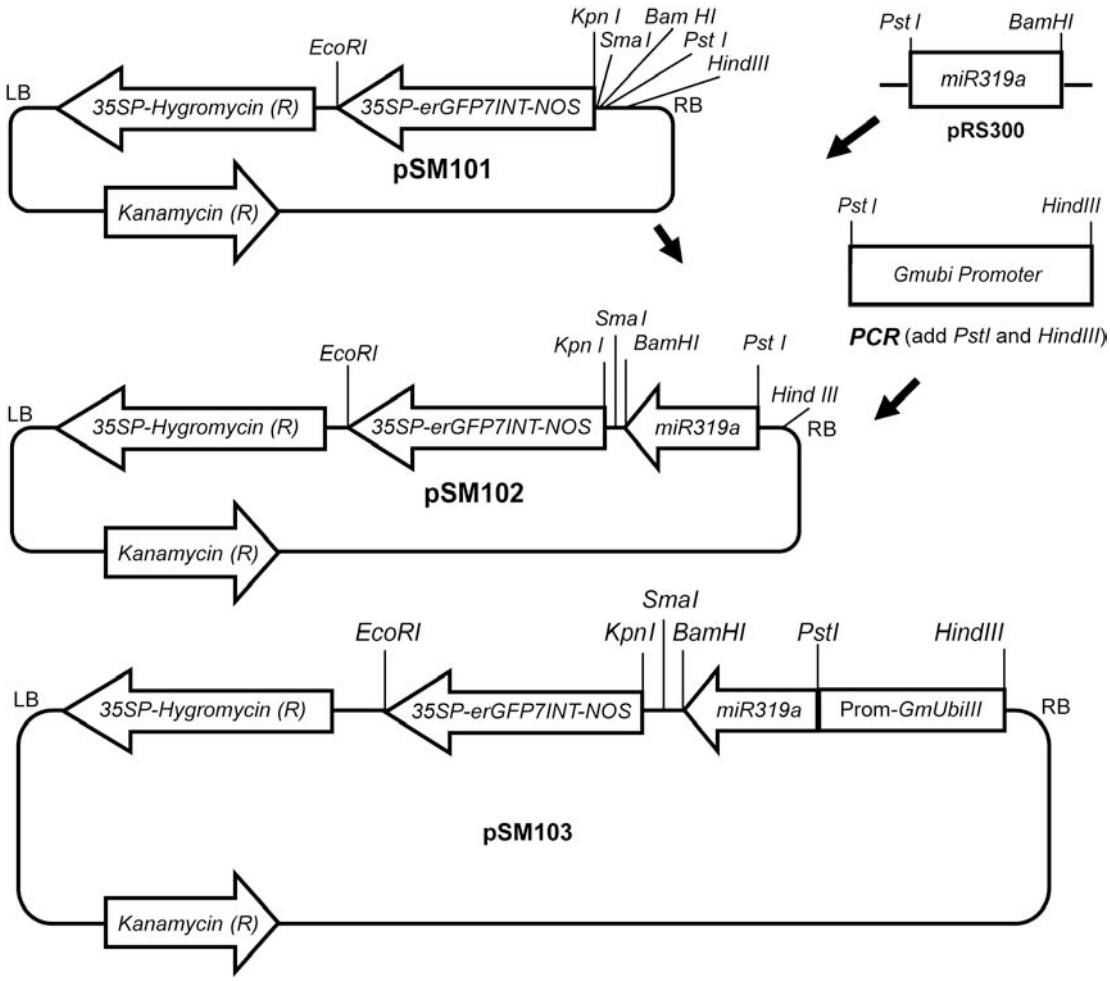


Figure 2.

Construction of pSM103, a plasmid for amiRNA-mediate gene silencing in legume roots.

pSM101 is an intron-GFP7 *Agrobacterium*-compatible binary plasmid for plant transformation, derived from pCAMBIA1300 by addition between T-DNA borders of a 35S promoter-erGFP7INT-NOS terminator (Mankin et al. 2001). pSM102 is pSM101 with a promoterless Arabidopsis miRNA *miR319a* gene. pSM103 is pSM102

with the miR319a driven by a promoter from the soybean ubiquitin-3 gene for strong expression in legume roots. After opening with PstI and BamHI to remove the miR319a, pSM103 can receive user-designed amiRNA gene silencing constructs from the widely used plasmid pRS300 (http://wmd3.weigelworld.org/; Schwab et al 2006), or other constructs.

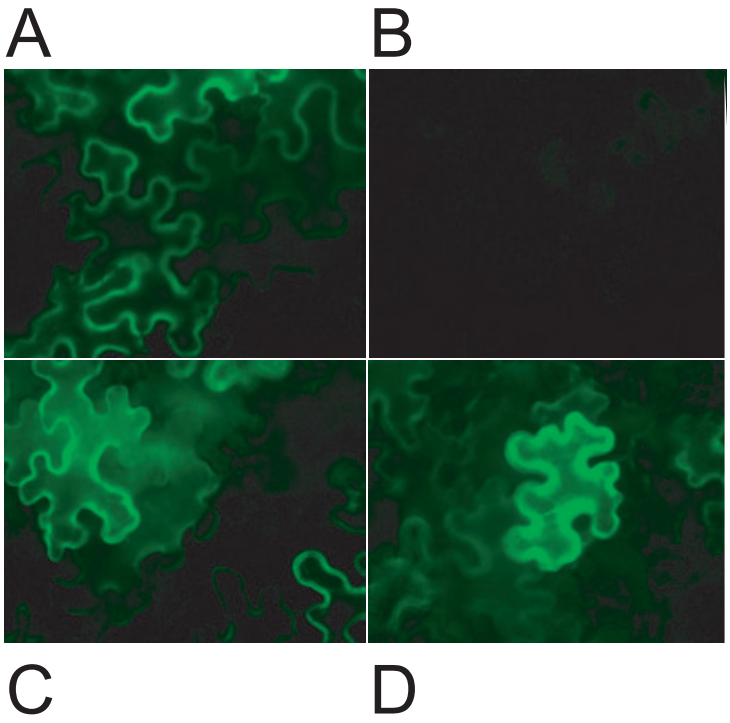


Figure 3.

Transient expression of pSM101 and pSM103 in *Nicotiana benthamiana* leaves.

A, B, C, The new binary vectors pSM101 and pSM103 were tested for GFP7 expression in *Nicotiana benthamiana* leaves. Bacterial suspensions were infiltrated through the abaxial surface of the leaf and GFP7 expression was monitored three days later by fluorescence microscopy. Leaves were infiltrated with: **A,** positive control *A. tumefaciens GV3101* pGWB1::35S-erGFP7INT-NOS; **B**, negative control

untransformed *A. tumefaciens GV3101;* **C** & **D**, *A. tumefaciens GV3101* pSM101 (C) or pSM103 (D) that carry intron GFP7 constructs.

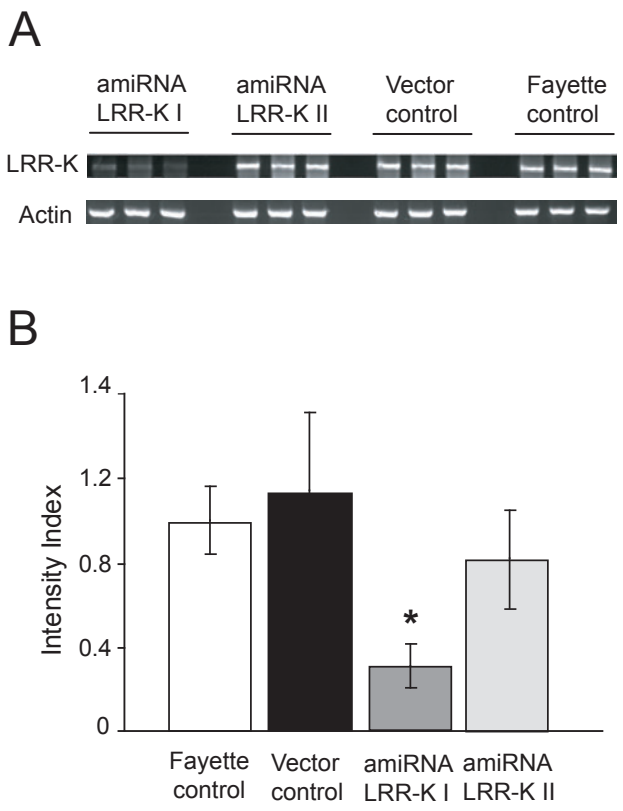
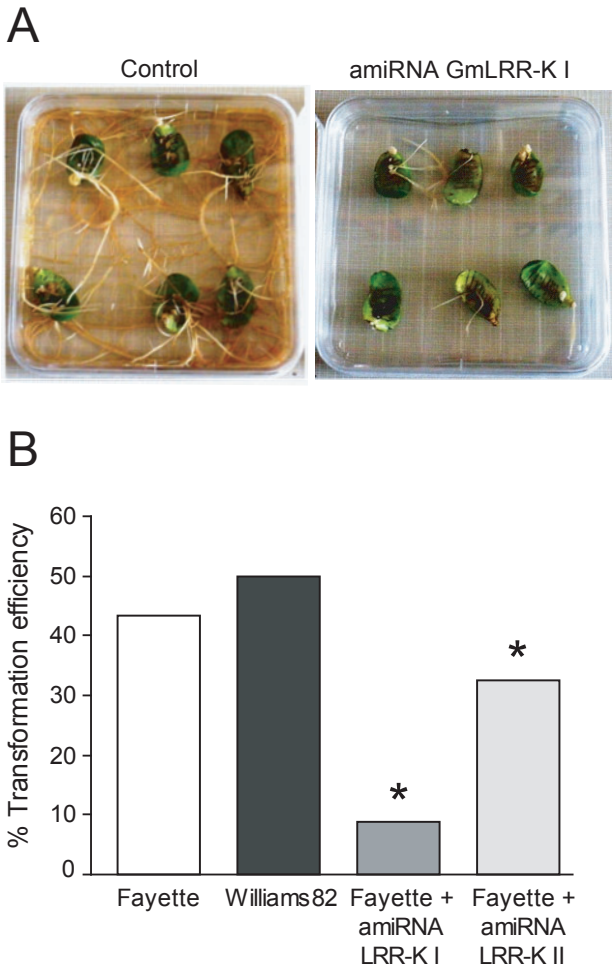


Figure 4.

amiRNA-mediated silencing of expression of the *Rhg1*-locus LRR-kinase gene.

A, Semi-quantitative RT-PCR to monitor expression level of the *Rhg1*-locus LRR-kinase gene Glyma18g02680.1/TC204550 (LRR-K) in Fayette transgenic roots generated using pSM103-derived plasmids expressing amiRNA GmLRR-K I or amiRNA GmLRR-K II, or GFP7-positive Fayette transgenic roots generated using the negative control vector pSM103 (vector control), or GFP-negative Fayette hairy roots. Actin expression was monitored in parallel from the same samples as a control. Results from three independent roots are shown for each construct; gels were stained with ethidium bromide. Similar results were observed in a second

experiment using independently derived roots. **B,** Relative band intensity of semi-quantitative RT-PCR products. Data from the experiment shown in part A and from a similar but independent transformation experiment are shown. Band intensities were normalized to the mean value for the three Fayette control roots from the same experiment. Mean ± std error of mean are shown; statistical analysis was performed on the normalized data. Roots expressing amiRNA GmLRR-K I exhibited a significantly reduced mRNA abundance for the *Rhg1*-locus LRR-kinase gene, as



indicated with the * (ANOVA; Tukey comparison of means; p = 0.0023).

Figure 5.

Root production by soybean cotyledons transformed using *A. rhizogenes* carrying amiRNA silencing constructs.

A, Fayette cotyledons three weeks after treatment with *A. rhizogenes* pSM103 (control, plasmid carries *A. thailiana* miRNA319a lacking any known soybean

target), or with *A. rhizogenes* pSM103 derivative carrying amiRNA GmLRR-K I. Representative explants are shown.

B, Transformation efficiency reduction when *A. rhizogenes* deliver amiRNAs targeting the *Rhg1*-locus LRR-kinase gene. Efficiency is the % of treated cotyledons that produced at least one GFP7-positive root segment; * indicates significant difference from Fayette control (χ^2-test, see text). Cotyledons transformed with pSM103 derivates carrying amiRNA GmLRR-K I developed roots approximately two weeks later than those transformed with the control pSM103 vector, and developed fewer roots per cotyledon. This effect was also present, but less dramatic in the cotyledons that received the amiRNA GmLRR-K II.

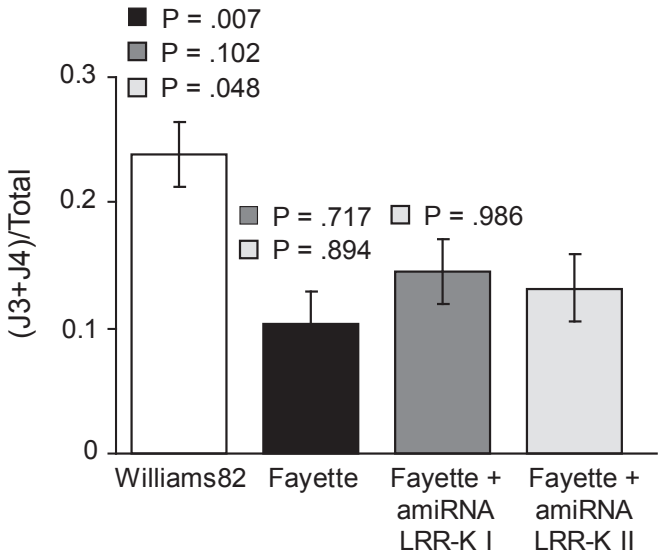


Figure 6.

Nematode demographic assay for SCN resistance in roots transformed with amiRNA silencing the *Rhg1*-locus LRR-kinase gene.

Ratio of total infecting nematode population that gained appearance of J3 and J4/adult female stages two weeks after inoculation onto *A. rhizogenes*-transformed roots. Williams 82 (susceptible) and Fayette (resistant) soybean tissue (left two bars) were transformed using *A. rhizogenes* carrying the pSM103 negative control plasmid; Fayette + amiRNA received the designated amiRNAs that target the *Rhg1*-locus LRR-kinase gene (amiRNA LRR-K I and amiRNA LRR-K II). Graph shows mean ± std error of mean from three independent nematode infection assays that used roots from three independent transformation experiments (overall n > 18 for each treatment). P values above each bar are color-coded for contrasts with the bar of the designated color (null hypothesis: the observed ratio between the two treatments was similar; tested by ANOVA using SAS Proc Mixed).

3. Copy number variation of multiple genes at *Rhg1* mediates nematode resistance in soybean

This chapter was previously published as:

David E. Cook*, Tong Geon Lee*, Xiaoli Guo*, Sara Melito, Kai Wang, Adam Bayless, Jianping Wang, Teresa J. Hughes, David K.Willis, Thomas Clemente, Brian W. Diers, Jiming Jiang, Matthew Hudson*, Andrew F. Bent*. 2012. Copy number variation of multiple genes at Rhg1 mediates nematode resistance in soybean. Science, 338(6111): 1206-1209.

*These authors contributed equally to this work

I was responsible for the results of gene silencing experiments for all genes except *Glyma18g02590*, qPCR transcript abundance estimates between soybean lines, providing materials for fiber-FISH and northern analysis, testing the impacts of overexpression on

nematode development and transcription in conjunction with Xiaoli Guo, along with designing many of the experiments, making the figures, and writing the paper with Andrew Bent.

3.1. Abstract

The *rhg1-b* allele of soybean is widely used for resistance against soybean cyst nematode (SCN), the most economically damaging pathogen of soybeans in the United States. Gene silencing showed that genes in a 31 kb segment at *rhg1-b*, encoding an amino acid transporter, an α-SNAP protein and a WI12 (wound-inducible domain) protein, each contribute to resistance. There is one copy of the 31 kb segment per haploid genome in susceptible varieties, but ten tandem copies in the *rhg1-b* haplotype. Overexpression of the individual genes in roots was ineffective, but overexpression of the genes together conferred enhanced SCN resistance. Hence *Rhg1*-mediated SCN resistance is conferred by copy number variation that increases the expression of a set of dissimilar genes in a repeated multi-gene segment.

One-Sentence Summary: Resistance to the most damaging disease of soybean is conferred by ten tandemly repeated copies of a group of dissimilar genes at the widely used *rhg1-b* locus.

3.2. Main Text

Soybean (*Glycine max*) is the world's most widely used legume crop, providing 68% of world protein meal as well as food oil and renewable fuels, and a farm gate value of over $35 billion in the U.S. alone (http://www.soystats.com/). Soybean cyst nematode (SCN; *Heterodera glycines*) is the most economically damaging pathogen of soybean, estimated to cause over $1 billion dollars in annual losses. SCN has infested most major soybean producing areas worldwide, and there are no practical means of eradication (Niblack et al., 2006).

SCN molts through multiple juvenile and adult life stages, including obligate endoparasitic stages on plant roots, to complete its life cycle (Niblack et al., 2006). Infective J2 juveniles invade roots of both susceptible and resistant soybean hosts, then reprogram host root cells to form feeding sites using highly-evolved, secreted nematode effectors (Davis et al., 2008; Gheysen and Mitchum, 2011).

The soybean *Rhg1* (Resistance to *H. glycines)* quantitative trait locus on chromosome 18 consistently contributes much more effective SCN resistance than any other known loci (Concibido et al., 2004; Kim et al., 2011). *Rhg1* disrupts the formation and/or maintenance of most potential nematode feeding sites (Niblack et al., 2006). Roughly 90% of the commercially cultivated soybean varieties marketed as SCN-

resistant in the central U.S. use the *rhg1-b* allele (haplotype), derived from the soybean line PI 88788, as the main SCN resistance locus. The molecular basis of this SCN-resistance has remained unclear.

Genetic mapping has placed *rhg1-b* in an interval that corresponds to a 67 kb segment carrying 11 predicted genes in the genome of the SCN-susceptible but fully sequenced Williams 82 soybean variety (Kim et al., 2010; Schmutz et al., 2010). It was recently suggested that an amino acid polymorphism in the *Glyma18g02590*-encoded α-SNAP protein in this interval contributes to SCN resistance (Matsye et al., 2012), although the authors indicated that this polymorphism does not account for *rhg1-b*-mediated resistance. None of the gene products within the *rhg1-b* genetic interval resemble canonical plant immune receptors (Dodds and Rathjen, 2010).

In the present study, genes from the *rhg1-b* interval (Kim et al., 2010) were silenced to test for impacts on SCN resistance (M.M., 2012). Transgenic soybean roots expressing artificial micro-RNA (amiRNA) or hairpin (RNAi) constructs were produced using *Agrobacterium rhizogenes* (Narayanan et al., 1999; Smith et al., 2000; Ossowski et al., 2008). Soybean resistance to SCN was measured two weeks after root inoculation by determining the proportion of the total nematode population that had advanced past the J2 stage in each root (Fig. 1A), relative to known resistant and susceptible controls (Melito et al., 2010). Silencing any one of three closely linked genes at the *rhg1-b* locus of the SCN-resistant soybean variety Fayette significantly reduced SCN resistance (Fig. 1B). Depletion of resistance was dependent on target transcript reduction (fig. S1). Silencing other genes in and around the locus did not impact SCN resistance (e.g., Fig 1B, genes

Glyma18g02570 and *-2620*) (M.M., 2012). The three *Rhg1* genes discovered to contribute to SCN resistance encode a predicted amino acid transporter (*Glyma18g02580*), an α-SNAP protein predicted to participate in disassembly of SNARE membrane trafficking complexes (*Glyma18g02590*), and a protein with a WI12 (wound-inducible protein 12) region but no functionally characterized domains (*Glyma18g02610*) (Yen et al., 1999; Jahn and Scheller, 2006; Okumoto and Pilot, 2011).

Concurrent study of the physical structure of the *rhg1-b* locus revealed an unusual genomic configuration. A 31.2 kb genome segment encoding the above-noted genes is present in multiple copies in SCN resistant lines (Figs. 2, 3). The DNA sequence of fosmid clone inserts carrying genomic DNA from the *rhg1-b* genetic interval identified a unique DNA junction, not present in the published Williams 82 soybean genome, in which the intergenic sequence downstream of (centromeric to) *Glyma18g02610* is immediately adjacent to a 3' fragment of *Glyma18g02570* (Fig. 2A). The genomic repeat contains full copies of *Glyma18g02580*, *-2590* , *-2600* and *-2610* as well as the final two exons of *Glyma18g02570*. Whole-genome shotgun sequencing of a line containing *rhg1-b* revealed ten-fold greater depth of coverage of this interval relative to surrounding or homeologous regions (Fig. 2B), suggesting the presence of multiple repeats. Further PCR and sequencing tests confirmed the presence of the *Glyma18g02610-2570* junction in DNA from multiple SCN-resistant soybean accessions, while the junction was not detected in four tested SCN-susceptible varieties including Williams 82 (Fig. 2C and fig. S2). The shared identity of the junction sites in disparate sources of SCN resistance suggests a shared origin of the initial resistance-conferring event at *Rhg1*.

Gene expression analysis using quantitative PCR (qPCR) determined that the three genes found to significantly impact SCN resistance show higher transcript levels in roots of SCN-resistant varieties than in susceptible lines (Fig. 2D and fig. S2). Full-length transcripts were confirmed for *Glyma18g02580*, *-2590* and *-2610,* no transcript was detected for *Glyma18g02600* and no hybrid repeat-junction transcript was detected for *Glyma18g02570* (fig. S2). The above suggested that elevated expression of one or more of the SCN-impacting genes could be a primary cause of elevated SCN resistance.

Fiber-FISH (fluorescence in situ hybridization) was used to directly view the arrangement and copy number of the 31 kb repeat segment in different haplotypes of the *Rhg1* locus. The hybridization pattern and DNA fiber length estimates (Fig. 3 and table S1) indicate a single copy of the repeat in Williams 82, as in the reference soybean genome (Schmutz et al., 2010). In Fayette, fiber-FISH revealed ten copies per DNA fiber of the repeat segment, in the same configuration throughout the multiple nuclei sampled, in a pattern consistent with ten direct repeats abutting head-to-tail (Fig. 3 and table S1). In samples from Peking, another common source of SCN resistance, three copies per DNA fiber were present in direct repeat orientation (Fig. 3). No additional copies (e.g., at other loci) were evident. *Rhg1* repeat copy number expansion is likely to have occurred by unequal exchange meiotic recombination events between homologous repeats.

Amino acid polymorphism or overexpression of any one of the three identified *rhg1-b* genes did not account for SCN resistance. From all available *rhg1-b* sequence reads (across multiple repeat copies), no predicted amino acid polymorphisms relative to Williams 82 were identified for *Glyma18g02580, Glyma18g02600* or *Glyma18g02610*.

Some copies of *Glyma18g02590* from *rhg1-b* resemble the Williams 82 sequence while others contain a set of polymorphisms, notably at the predicted C-terminal six amino acids of the predicted α-SNAP protein (table S2, confirmed by cDNA sequencing). However, expressing this polymorphic *rhg1-b*-type *Glyma18g02590* downstream of a strong constitutive promoter or native promoter sequence did not increase the SCN resistance reaction of Williams 82 transgenic roots (Fig. 4 and fig. S3), suggesting that *rhg1-b* SCN resistance requires more than this 2590 amino acid polymorphism. Overexpression of *Glyma18g02580* or *Glyma18g02610* also failed to increase SCN resistance (Fig. 4).

Given the above, simultaneous overexpression of genes within the 31 kb repeat segment was tested as a possible source of SCN resistance. In two separate experiments using >25 independent transgenic events for each DNA construct, resistance to SCN was significantly increased in SCN-susceptible Williams 82 by simultaneous overexpression of the set of genes (Fig. 4; see also fig. S4A). A DNA construct overexpressing *Glyma18g02580*, *-2590* and *-2610* but not *-2600* also conferred enhanced SCN resistance. The collected findings indicate *Rhg1*-mediated SCN resistance is attributable to elevated expression of *Glyma18g02580*, *-2590* and *-2610*.

These results reveal a novel mechanism for disease resistance: an expression polymorphism for multiple disparate but tightly linked genes, derived through copy number variation at the *Rhg1* locus. This suggests future approaches to enhance *Rhg1*-mediated quantitative resistance against the globally important SCN disease of soybean, for example through isolation of soybean lines that carry more copies of the 31 kb *Rhg1*

repeat. Transgenic overexpression of the native or altered genes may improve SCN resistance, and/or be applicable in other species for resistance to other endoparasitic nematodes.

The biochemical mechanisms of *Rhg1*-mediated resistance remain unknown. Other sequenced plant genomes do not carry close homologs of the predicted Glyma18g02610 protein, although a wound-inducible protein in ice plant with 55% identity has been studied (Yen et al., 1999). Modeling of the Glyma18g02610 predicted tertiary structure suggested (M.M., 2012) that Glyma18g02610 may participate in the production of phenazine-like compounds that are toxic to nematodes. The Glyma18g02590 α-SNAP protein is likely involved in vesicle trafficking and may influence exocytosis of products that alter feeding site development or nematode physiology (Frei dit Frey and Robatzek, 2009). Because it is one of at least five α-SNAP homologs encoded in the soybean reference genome, Glyma18g02590 may have undergone subfunctionalization or neofunctionalization (Conrad and Antonarakis, 2007). The Glyma18g02580 protein and its most closely related plant transporter proteins are not functionally well characterized, but Glyma18g02580 contains a predicted tryptophan/tyrosine permease family domain. Tryptophan shares structural similarity with and is a precursor of the auxin hormone indole-3-acetic acid, suggesting the intriguing possibility that Glyma18g02580 may impact functionally important auxin levels or distribution (Davis et al., 2008; Gheysen and Mitchum, 2011). Together, these genes apparently create an unfavorable extracellular environment at nematode feeding sites.

Growing evidence from metazoa and plants suggests that genome structural variation is a frequent and powerful driver of phenotypic diversity (Sebat et al., 2004; Springer et al., 2009). Copy number variation of chromosomal subsegments (beyond simple duplication) can impact gene expression levels (Stranger et al., 2007), and single gene copy number variation contributes to a number of adaptive traits in humans, plants and insects (Perry et al., 2007; Schmidt et al., 2010; Pearce et al., 2011; Wingen et al., 2012). Recent analyses of genome-architecture in sorghum, rice, and soybean have reported high levels of copy number variation, and a tendency for these genomic regions to overlap with postulated biotic and abiotic stress-related genes (Yu et al., 2011; Zheng et al., 2011; McHale et al., 2012)

The present work provides a concrete example of copy number variation in which the repeat encodes multiple gene products that contribute to a valuable disease resistance trait. Single-copy clusters of genes that are functionally related but non-homologous are highly unusual in multicellular eukaryotes, but such clusters have been reported in association with plant secondary metabolism (Frey et al., 1997; DellaPenna and O'Connor, 2012). Given the repetitive and plastic nature of plant genomes and the relatively underexplored association between copy number variation and phenotypes, it seems likely that a number of other complex traits are controlled by this type of structural variation.

3.3. Acknowledgements

This work was supported by the United Soybean Board (A.F.B.), USDA CSREES (M.H.), Illinois Soybean Association (M.H. and B.W.D.), National Science Foundation grant DBI-0922703 (J.J.), Wisconsin Experiment Station Hatch Award (A.F.B.), and the Pioneer Fellowship in Plant Pathology awarded to D.E.C. by the American Phytopathological Society through a gift from Pioneer Hi-Bred. We thank A.E. MacGuidwin for suggestions regarding SCN experiments and J. Palmer for assistance with RNA blots.

3.4. Materials and Methods

3.4.1. *Agrobacterium rhizogenes* Soybean Root Transformation

A. rhizogenes strain Arqua1 was transformed by freeze-thaw as previously reported (Hofgen and Willmitzer, 1988; Wise et al., 2006). The cells were plated on selective media with the appropriate antibiotic and incubated at 28 °C for two days. *A. rhizogenes* strain Arqua1 was received from Dr. Jean-Michel Ane, University of Wisconsin Madison. Soybean seeds lacking macroscopic signs of fungal or viral contamination were surface-sterilized for 16-20 h in a desiccator jar with chlorine gas generated by adding 3.5 ml 12 N HCl into 100 ml household bleach (6% sodium hypochlorite). At least 20 seeds per experiment were plated onto germination media (Gamborg's B5 salts (3.1g/L), 2% sucrose, 1× Gamborg's B5 vitamins, 7% Noble agar, pH 5.8) in 100 × 25 mm Petri plates. Plates were wrapped with Micropore tape (3M, St. Paul, MN) and incubated at 26°C in a growth chamber (18/6 light/dark hours) for approximately one week. Soybean cotyledons were harvested 5-7 days after germination by gently removing them from the hypocotyls with sterile forceps. With a sterile forceps and Falcon #15 scalpel, several shallow slices were made across the abaxial surface of the cotyledons after dipping the scalpel in *A. rhizogenes* suspension (OD_{600} 0.6 - 0.7 in sterile ddH_20). The cotyledons were then placed abaxial-side down on a co-culture medium (CCM) (0.31 g/L Gamborg's

B5 salts, 3% sucrose, 1× Gamborg's B5 vitamins (BioWorld, Dublin OH), 0.4 g/L L-cysteine, 0.154 g/L dithiothreitol, 0.245 g/L sodium thiosulfate, 40 mg/L acetosyringone, 5% Noble agar, pH 5.4) in 100 × 15 mm Petri plates with a piece of 70 mm filter paper (Whatman, Piscataway, NJ) on the surface of the agar to prevent *A. rhizogenes* from overgrowing. Plates were wrapped with parafilm and incubated in the dark at room temperature for three days. The explants were then transferred to a hairy root medium (HRM) of 4.3 g/L MS salts (Sigma Co., St. Louis, MO), 2% sucrose, 1× Gamborg's B5 vitamins (BioWorld, Dublin, OH), 7% Noble agar, 0.15 g/L cefotaxime, 0.15 g/L carbenicillin, pH 5.6 in 100 × 15 mm Petri plates, wounded side up. Plates were wrapped with Micropore tape and incubated in the dark at room temperature until roots emerged, usually in around 2 weeks. Transgenic soybean roots were detected based on plasmid vector-encoded GFP expression, using a fluorescence stereomicroscope (LEICA MZ FL III with GFP2 filter). Transgenic soybean root tip segments (2-3 cm) were transferred to HRM. Roots that were expressing incomplete strips of fluorescence (chimeras) or exhibiting overall low levels of GFP fluorescence were avoided. Independent transgenic events, generated from different inoculation sites or different cotyledons, were maintained separately for RNA extraction and nematode demographic assays.

3.4.2. Nematode maintenance

An SCN population from Racine, Wisconsin (Hg-type 7), collected by Ann MacGuidwin (University of Wisconsin-Madison), was maintained on the susceptible soybean cultivar Williams 82. Seeds were germinated between two damp pieces of paper towel that were

rolled-up and placed vertically in a glass beaker with a small amount of water at the bottom for 2-4 days. Germinated seeds were then planted in autoclaved 4:1 sand : soil mixture and inoculated with 2000 eggs of *H. glycines* per plant, and grown in a 28 °C growth chamber. Cysts were collected ~50 days after infection when soybeans were at R2 (full flowering) and extracted from soil and roots using sieves and centrifugation. Briefly, soil and roots from infected pots was placed in a pitcher of water and agitated. The soil-cyst-water slurry was passed over a 710 µm - 250 µm sieve tower, and the mixture from the 250 µm sieve was backwashed into a 50 mL plastic conical tube. The tubes were centrifuged at 2 000 rpm for 4 minutes then the supernatant was poured off. A 60% sucrose solution was added to the tubes, stirred, and centrifuged at 2000 rpm for 2 min. Cysts in the supernatant were then collected over a 250 µm sieve. Collected cysts were stored at 4°C in sealable plastic bags containing twice-sterilized flint sand.

3.4.3. Nematode demographics assay

Nematode demographics assays were performed as in (Melito et al., 2010). Vigorous new root segments (2-3 cm including root tip) were utilized. All roots (all genotypes within an experiment) were coded with a random number prior to inoculation, to mask root genotype information from the investigators who stained roots two weeks later and determined the number of nematodes in each nematode development category. For inoculum, *H. glycines* eggs were collected by breaking open cysts with a large rubber stopper and collecting the eggs on a sieve stack consisting of 250 µm - 75 µm - 25 µm sieves (USA Standard Testing Sieve). Eggs were collected from the 25 µm sieve and

rinsed. Eggs were placed in a hatch chamber (Wong et al., 1993) with 3 mM $ZnCl_2$ for hatching at room temperature in the dark for 5-6 days. Hatched J2 nematodes were surface-sterilized for 3 min in 0.001% mercuric chloride and washed three times with sterile distilled water, then suspended in room temperature 0.05% low-melting point agarose to facilitate even distribution (Baum et al., 2000). The number of active nematodes was determined by viewing an aliquot under a stereomicroscope at least one-half hour after surface-sterilization and washing, and 200-250 active J2s were inoculated onto each fresh root segment. Inoculated roots with nematodes were maintained on HRM media at 28 °C; substantial root growth typically occurred during the subsequent two weeks. Nematode infection and development within these root systems was monitored by clearing and staining with acid fuchsin (Byrd et al., 1983), typically 15 days post inoculation (dpi). The nematode demographic assay was then completed by recording the number of nematodes in each root system that exhibited a morphology resembling either J2 (thin), J3 (sausage-shaped), elongated male, or J4/adult female nematodes, as noted in text and figures. Typically, 20-80 nematodes were present in each root; roots containing fewer than ten nematodes were excluded from further analysis. Results were expressed as % of nematodes that had developed beyond J2 stage ([J3 + adult males +adult females]/[J2 + J3 + adult males +adult females]). Each data point was normalized to the mean for Williams 82 roots transformed with empty vector, from the same experiment. All reported data are based on at least two independent biological replicate experiments ($n > 12$ independently transformed roots for each bar on a bar graph). A total of 13 genes within and adjacent to the *rhg1-b* genetic interval defined by Kim et al. (Kim et al., 2010)

were tested in at least one set of nematode demographics assays with at least one gene silencing construct. All genes that gave promising assay results were retested at least one additional time. No reproducible impacts on SCN resistance were observed other than for the three genes shown in Fig. 1 (*Glyma18g02580*, *Glyma18g02590* and *Glyma18g02610*).

3.4.4. Vector Construction for Soybean transformation

Binary vectors pSM101 and pSM103 for soybean transformation were constructed as previously described (Melito et al., 2010). To generate and clone soybean amiRNAs, the Web microRNA Designer (http://wmd3.weigelworld.org) and protocols were used. The concept is more thoroughly documented in other references (Schwab et al., 2006; Ossowski et al., 2008). Soybean DNA was extracted from either expanding soybean trifoliates or soybean roots using a previously reported CTAB method (Doyle and Dickson, 1987) . PCR fragments for amiRNA construction were TA cloned using pCR8/GW/TOPO TA cloning kit (Life Technologies Corp., Carlsbad CA) (Table S3 13-24). Binary vectors pGRNAi1 and pGRNAi2 for soybean transformation were a gift from Wayne Parrot, University of Georgia (unpublished). For each hairpin, a 300-600 bp DNA fragment was PCR amplified (Table S3 1-12) using Phusion HF polymerase (New England Biolabs, Ipswich, MA) and iScript cDNA synthesis kit (Biorad, Hercules, CA) as a template, as per manufacturer's instructions. PCR products were TA cloned as previously described. Primers used to generate the DNA fragments were designed to contain restriction sites *Avr* II/*Asc* I (forward primer) and *Bam*H I/*Swa* I (reverse primer)

to allow cloning into pGRNAi1 and pGRNAi2. To generate the first arm of the hairpin, the insert and vector were sequentially digested with restriction endonucleases *Swa* I and *Asc* I using manufacturer's recommended protocol (New England Biolabs, Ipswich, MA). DNA was separated on a 1.0% agarose gel stained with ethidium bromide, and respective DNA fragments were gel purified using Qiaquick gel extraction kit (Qiagen, Valencia, CA) and ligated together overnight at 4°C using T4 DNA ligase (Promega, Madison, WI). The same procedure was used to insert the second arm of the hairpin construct using the restriction endonucleases *Bam*H I and *Avr* II. To construct single gene overexpression vectors for *Glyma18g02580* (Table S3 70, 71), *Glyma18g02590* (Table S3 57, 58) and *Glyma18g02610* (Table S3 55, 56), full-length ORFs were PCR amplified from cDNA of Fayette using Phusion HF polymerase and TA cloned in pCR8/GW/TOPO as previously described. *Glyma18g02600* (Table S3 67, 68) was cloned from genomic DNA by similar methods, as no *Glyma18g02600* cDNA could be detected in root cDNA libraries. The *Glyma18g02610* and *Glyma18g02590* ORFs were recombined with pGWB14 (CaMV 35S promoter, 3× HA-NOS terminator) (Nakagawa et al., 2007) using LR clonase reaction (Life Technologies Corp., Carlsbad, CA) per manufactures instructions. *Glyma18g02610* (Table S3 55, 59) was PCR amplified from pGWB14 and TA cloned into pCR8. This vector and pSM103 were digested with *Xba* I/*Kpn* I and ligated to yield $GmUbi_{prom}$*:2610-HA:*NOS_{term} (OE:2610-HA). The same procedure was used for *Glyma18g02590* (Table S3 57, 59), except the amplicon contained *Xba* I/*Sal* I sites and was TA cloned into pCR8. *2590-HA:*NOS_{term} and pSM103 were digested with *Xba* I/*Sal* I and ligated to yield $GmUbi_{prom}$*:2590-HA:*NOS_{term} (OE:2590-HA). The full OE:2590-HA

was also digested (*Xba* I/*Sa*l I) and ligated into pSM103 containing OE:2610-HA to yield OE:2610-OE:2590. To generate the four gene overexpression construct, the restriction sites *Pac* I, *Psp*OM I, and *Asc* I were added to pSM101 between sites *Pst* I/*Hind* II by annealing oligos (Table S3 62, 63) to generate pSM101+. The two gene overexpression cassette (OE:2610- OE:2590) was moved to the new pSM101+ using the restriction enzymes *Pst* I/*Kpn* I and ligation. A Nos promoter was added to *Glyma18g02600* in pCR8 using overlap PCR (Table S3 65-68) and TA cloned into pCR8. This vector was recombined with pGWB16 (no promoter, 4xMyc-NOS terminator) (Nakagawa et al., 2007) in an LR clonase reaction to yield Nos_{prom}*:2600-myc:*Nos_{term} (OE:2600-myc). OE:2600-myc was PCR amplified (Table S3 66, 72) and TA cloned into pCR8, and subcloned into pSM101+ (OE:2610-OE:2590) using restriction enzymes *Hind* III/*Asc* I to yield the three gene overexpression vector (OE:2610-OE:2590-OE:2600). A Nos promoter was added to *Glyma18g02580* in pCR8 using overlap PCR with primers 71-74 and TA cloned into pCR8. This vector was used with pGWB16 in an LR clonase reaction to yield Nos_{prom}*:2580-myc:*Nos_{term} (OE:2580-myc). OE:2580-myc was amplified (Table S3 72, 75) and TA cloned, then subcloned into the three gene overexpression vector resulting in the four gene overexpression vector pSM101+ OE:2610-OE:2590-OE:2600-OE:2580. The native Fayette *Glyma18g02590* (2590_{FayP}*:*2590_{Fay}) construct for Williams 82 complementation was subcloned from a fosmid containing the desired allele. A 6.5 kb DNA fragment containing the PI 88788 *Glyma18g02590* was isolated from a fosmid following *Sal* I digestion and cloned into pSM101 using the *Sal* I restriction site. This sequence contained approximately 1 kb of 5' regulatory DNA sequence. An additional

600 bp of 5' regulatory sequence directly upstream of the subcloned region was added to the construct by amplifying a PCR product (Table S3 79, 80) from the fosmid and inserted using the restriction enzymes *Hind* III/*Sal* I. The resulting construct contained approximately 1.6 kb of naturally occurring 5' regulatory sequence of the Fayette *Glyma18g02590* allele. Vector sequences were confirmed at various steps using Sanger sequencing with ABI Big Dye cycle sequencing kit (dideoxy chain-termination) and ABI 3730xl DNA Analyzers (Life Technologies Corp., Carlsbad, CA), using the DNA sequencing service at the University of Wisconsin-Madison Biotechnology Center.

3.4.5. Quantitative Real Time PCR

Quantitative PCR (qPCR) was performed using either the MyIQ or CFX96 real-time PCR detection system (BioRad, Hercules, CA). cDNA was synthesized from RNA using iScript cDNA synthesis kit (Biorad, Hercules, CA) per manufactures protocol by adding 0.825 µg to 1.0 µg of RNA depending on the experiment. Total RNA was extracted from root tissue of conventional and transgenic soybeans. RNA was extracted from conventional soybean plants grown in Metro mix for two weeks at 26°C and 16 hours light prior to tissue collection. Roughly 200 mg of tissue was collected from each plant, immediately flash-frozen in liquid nitrogen and stored at -80°C. Transgenic root material was collected from roots actively growing on HRM as previously described. Roughly 50-100 mg of tissue was collected from each root, flash frozen in liquid nitrogen and stored at -80°C. RNA was extracted using either the RNeasy Mini Kit (Qiagen, Valencia, CA) or TRIzol reagent (Life Technologies Corp., Carlsbad, CA) following manufactures

protocols. RNA concentrations were determined using the NanoDrop-1000 spectophotomoter (Thermo Scientific, Waltham, MA). DNA was removed from RNA samples using either RNase-free DNase I (Qiagen, Valencia, CA) or DNA-free (Life Technologies Corp., Carlsbad, CA) following manufacture protocols. RNA integrity was determined using the 2100 BioAnlyzer (Agilent Technologies, Santa Clara, CA) or 500 ng of total RNA was run on a 1.2% agarose gel stained with ethidium bromide and visualized under UV-light to ensure RNA quality following extraction. qPCR reactions were carried out using either IQ SYBR Green Supermix or SsoFast EvaGreen Supermix (Biorad, Hercules, CA). Primer concentrations for all reactions were between 0.2 µM and 0.3 µM. Two technical replicates were run per RNA. Efficiency curves were generated for qPCR primer pairs using cDNA from the cultivar Fayette or Williams 82 following a 3-4 step, 3-5 fold dilution. Following amplification, a melt curve program was performed. To ensure qPCR fluorescent signal was not the results of DNA, 100 ng of RNA extraction was added directly to IQ SYBR Green Supermix or SsoFast EvaGreen Supermix with primers. DNA contamination was considered negligible if CT values were not detected until after 32-35 cycles. A control reaction was run in parallel using a known cDNA sample. Transcript abundance for genes at *Rhg1* was measured using primers 25-38. A total of six primer pairs were tested as reference genes (*EF1B, SKIP16, UNK2, ACT11, UNK1, TIP41*) (Table S3 39-50) (Hu et al., 2009) . Reference genes were validated using Bestkeeper anlysis (Pfaffl et al., 2004) Primer pairs *SKP16* and *TIP41* were selected and used in subsequent experiments. Transgenic roots expressing empty-vector constructs analogous to the vectors carrying gene silencing or gene expression

constructs were included in the experiments as controls and used to standardize gene expression. Results were considered to be at the limits of detection if CT values were >35 (i.e., for *Glyma18g02600* transcripts).

3.4.6. DNA repeat junction analysis

The presence of a repeat junction was confirmed using PCR (Table S3 81, 82) and soybean genomic DNA from SCN resistant cultivars Fayette, Hartwig, Newton and SCN susceptible cultivars Williams 82, Essex, Thorne and Sturdy. DNA extraction and PCR were performed as previously described. Possible impacts of retrotransposons on *Rhg1* locus evolution were investigated by searching for sequences with similarity to known plant retrotransposons. A 226 bp sequence with 73% identity to the 5' and 3' long terminal repeat (LTR) regions of Ty1/copia-like retrotransposons RTvr1 and RTvr2 is present 200 bp away from the *rhg1-b* duplication junction.

3.4.7. Statistical analysis

Data were analyzed by ANOVA using Minitab (v.14) with the General Linear Model and Tukey Simultaneous Test.

3.4.8. Fosmid clone sequencing and assembly

Seed of soybean Plant Introduction (PI) 88788 was obtained from the USDA soybean germplasm collection. Plants were grown in a growth chamber set at a photocyle of 18/6 hr (day/night), 23/20 °C (day/night), and 50% relative humidity for 1-2 weeks. Young

leaf tissue was collected from six to 15 individuals for each line. Genomic DNA was extracted using cetrimonium bromide (CTAB). Plant samples were ground to fine powder in liquid nitrogen, transferred to 20 ml of CTAB extraction buffer (2% CTAB, 100 mM Tris pH 9.5, 1.4 M NaCl, 1% PEG 6000, 20 mM EDTA, 2% polyvinylpyrrolidone, 2.5% β-mercaptoethanol), and placed at 65 °C for 1 hr. After incubation, an equal volume of Phenol:Chloroform:Isoamyl Alcohol (25:24:1, pH 6.7) was added to the tube, then centrifuged at 8,000 g at 10 °C for 10 min. The aqueous (top) phase was transferred to a new tube and an equal volume of chloroform:isoamyl alcohol (24:1) was added to the aqueous phase and centrifuged. The aqueous (top) phase was then transferred to a new tube and 0.7 volumes of isopropyl alcohol was added to the aqueous phase. After mixing well, the aqueous phase was centrifuged and the pellet resuspended in 70% EtOH, centrifuged at 7,500 g for 10 min. After centrifugation, the pellet was resuspended in 100 µl of TE (10 mM Tris pH 7.5, 1 mM EDTA). The DNA was treated with RNase A by incubating in 20 µg/ml RNase A at 37 °C for 1 hr. The PI 88788 fosmid library was constructed using the CopyControl™ Fosmid Library Production Kit (Epicentre, Madison, WI) following the manufacturer's protocol. Briefly, 20 µg of the size-fractionated DNA was used for end-repair. 35-45 kb fragment pools of DNA were cloned in the pCC1FOS™ Vector. Ligated DNA was packaged using the MaxPlax™ Lambda Packaging Extracts and transformed into the Phage T1-Resistant EPI 300 ™-T1^{R} *Escherichia coli* strain.

3.4.9. Whole-genome sequencing and read depth in duplicated region

Whole-genome shotgun sequencing of a soybean breeding line LD09-15087a, a near-isogenic line (NIL) that harbors *rhg1-b* from PI 88788, was conducted using Illumina technology. 1.5 μg of genomic DNA was sequenced using the Illumina HiSeq 2000 instrument with 100 bp paired-end sequencing at the University of Illinois Biotechnology Center. The DNA fragment size for the soybean whole-genome shotgun sequencing library was 600 bp; the library was loaded onto one lane of a flow cell and sequenced using version 3 of sequencing kits and Casava 1.8 (pipeline 1.9). 312,909,668 reads (about 28 × coverage of the 1.1 gb soybean genome) were generated with all positions having average quality scores 30 or higher. To examine the depth of the coverage within the duplicated region, reads from the sequencing were aligned to the Glyma1 version of the soybean genome assembly. Novoalign (v 2.08.01) (http://www.novocraft.com) with paired end options (PE 600,120) was used to align the reads to the reference genome. Approximately 95.1% of reads were aligned to the reference sequence. The number of reads aligned to the target interval was counted from a BAM file using SAMtools (v 0.1.18). Target interval is as follows: "Block" in Fig. 2B: a 31.2 kb region (1,632,225-1,663,455 on chromosome 18), "Block-1": the same size region as region of interest upstream, and "Block+1": the same size region as region of interest downstream. Homeologous regions on chromosome 11 ("Block" in Fig. 2B: 37,392,345-37,434,356 bp) and 2 ("Block": 47,772,323-47,791,521 bp) were identified using BLASTN.

3.4.10. Fiber-FISH

Soybean nuclei were lysed to release large chromosomal segments and, in contrast to more standard FISH methods, the chromosome segments were decondensed to generate extended DNA fibers before fixing to microscope slides and hybridizing to fluorescently labeled DNA probes. Young leaf tissues were collected from fast growing plants of Williams 82, Peking, and Fayette. Nuclei isolation, DNA fiber preparation, and fiber-FISH were performed following published protocols (Jackson et al., 1998). A fosmid clone spanning an *rhg1-b* repeat from PI 88788 was digested using the exonuclease SmaI (New England Biolabs, Ipswich, MA). The products of the restriction digestion were separated in a 0.7% gel and isolated using the Qiaex II gel extraction kit (Qiagen, Valencia, CA). DNA probes were labeled with either biotin-16-UTP or digoxigenin-11-dUTP (Roche Diagnostics, Indianapolis, IN) using a standard nick translation reaction. The fiber-FISH images were processed with Meta Imaging Series 7.5 software. The final contrast of the images was processed using Adobe Photoshop CS3 software. The cytological measurements of the fiber-FISH signals were converted into kilobases using a 3.21 kb/μm conversion rate (Cheng et al., 2002).

3.4.11. Transcript analysis

To confirm the annotation of transcripts at *Rhg1*(Schmutz et al., 2010), rapid amplification of cDNA ends (RACE) PCR was performed for *Glyma18g02580* (Table S3 95), *Glyma18g02590* (Table S3 87- 90) and *Glyma18g02610* (Table S3 91- 94) using the SMARTer RACE cDNA kit per manufacturer protocols (ClonTech, Mountain View,

CA). Following RACE, PCR products were TA cloned into pCR8/GW/TOPO as previously mentioned. Randomly chosen colonies were sequenced (Table S3 76, 77) as described to confirm the 5' and 3' ends of individual transcripts. Transcripts corresponding to *Glyma18g02600* could not be amplified from root cDNA, therefore we could not confirm its sequence. To detect potential transcript isoforms, northern analysis was conducted using standard methods (Sambrook and Russell, 2001). Probes were generated for *Glyma18g02570* (Table S3 83, 84). Absence of truncated *Glyma18g02570* transcripts (Table S3 85, 86) derived from 31.2 kb repeat junctions was also confirmed by PCR from cDNA, using a 2570 reverse primer and a forward primer in the most strongly predicted exon upstream of the repeat junction (Hebsgaard et al., 1996). Transcript abundance studies using qPCR also indicated that there is not a *Glyma18g02570*-like transcript produced by the repeated DNA insertion. *Glyma18g02570* transcript abundance was measured using primers (Table S3 25, 26) that amplify the final two exons and hence should amplify both the reference genome (full-length; Williams 82-like) *Glyma18g02570* transcript and possible hybrid *Glyma18g02570* transcripts that are transcribed from DNA that spans the repeat junction. If the repeated DNA produced an alternative transcript, these primers would amplify additional product from genotypes with the repeat. However, no differences in transcript abundance were detected between SCN-resistant vs SCN-susceptible varieties using *Glyma18g02570* primers 25 and 26.

3.4.12. Protein structure prediction and comparison

The protein structure for the predicted Glyma18g02610 gene product was modeled and proteins with the most homologous structures were identified using the default settings of Phyre2 (Kelley and Sternberg, 2009). The top scoring model identified, with 98% confidence, similarity of 48% of Glyma18g02610 to the PhzA/B subfamily of Delta(5)-3-ketosteroid isomerase/nuclear transport factor 2 family proteins. Other high-scoring models gave similar results, suggesting that Glyma18g02610 may participate in the production of phenazine-like compounds or other polycyclic compounds.

3.5. Figures

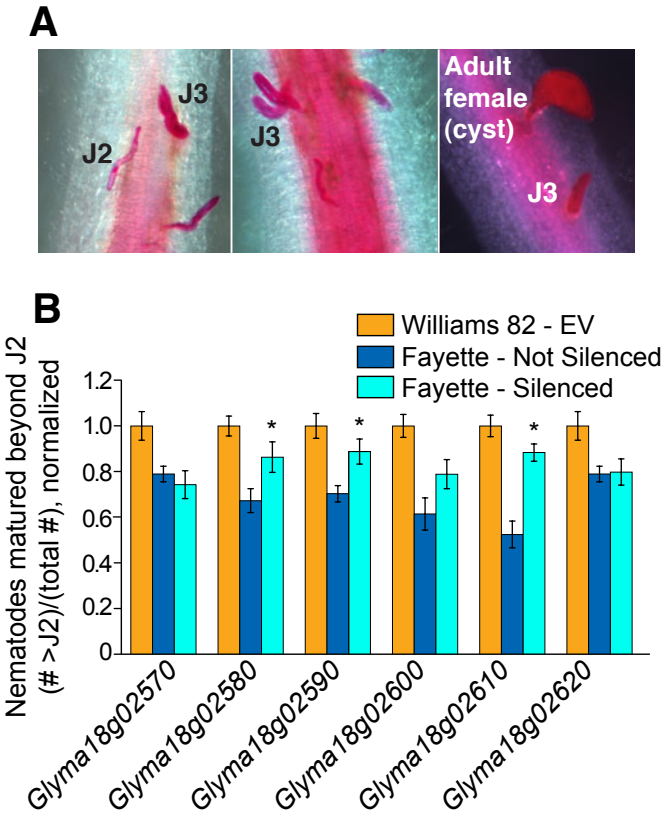

A
J3
J2
J3
Adult
female
(cyst)
J3
B
Williams 82 - EV
Fayette - Not Silenced
Fayette - Silenced
Nematodes matured beyond J2
(# >J2)/(total #), normalized
1.2
1.0
0.8
0.6
0.4
0.2
0
*
*
*
Glyma18g02570
Glyma18g02580
Glyma18g02590
Glyma18g02600
Glyma18g02610
Glyma18g02620

Figure. 1.

Three genes at *rhg1-b* contribute to SCN resistance. **(A)** Representative SCN-infested roots; root vascular cylinder and nematodes stained with acid fuchsin. Fewer nematodes progress from J2 to J3, J4, adult male or egg-filled adult female (cyst) stages in SCN-resistant roots. **(B)** SCN development beyond J2 stage in transgenic roots of soybean variety Fayette with the designated gene silenced, relative to Williams 82 (SCN-susceptible) and non-silenced Fayette (SCN-resistant) controls. Mean ± std. error of mean. *: Fayette (silenced) significantly different from Fayette (not silenced) based on ANOVA Tukey test, $p < 0.05$; $p > 0.1$ for *Glyma18g02600*. EV: transformed with empty vector.

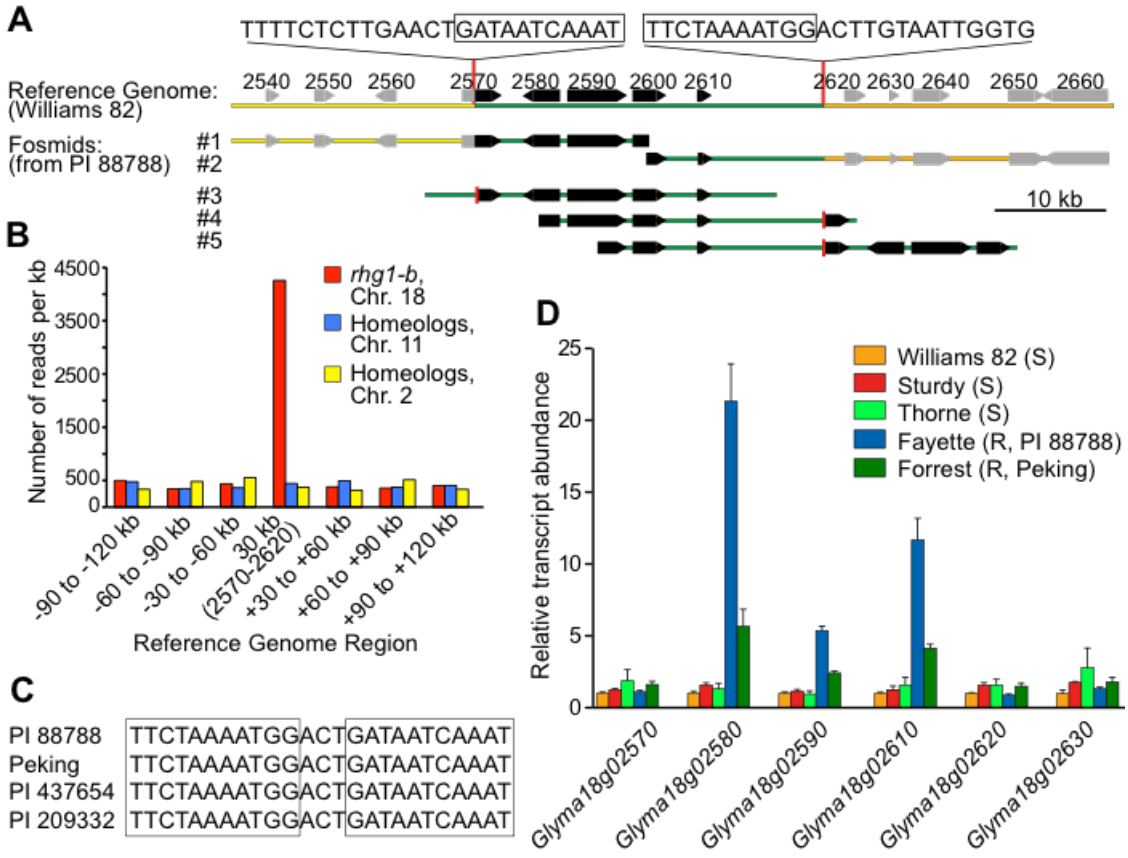


Figure. 2.

A 31.2 kb repeat that elevates expression of the encoded genes is present in SCN-resistant haplotypes of the *Rhg1* locus. **(A)** Schematic of *Rhg1* locus of Williams 82 (top), and five fosmid inserts from *rhg1-b* haplotype. Numbers and block icons refer to soybean genes (e.g., *Glyma18g02540*). Fosmids #3, 4 and 5 carry *rhg1-b* genome segments that span repeat junctions. **(B)** Sequence of unique *Rhg1* repeat junction not found in reference genome, from four different sources of SCN resistance. **(C)** Number of whole-genome shotgun sequencing reads corresponding to *Rhg1* (green region shown in (A)) was ten-fold greater than for adjacent genomic regions on chromosome 18 or for homeologous regions. **(D)** Transcript abundance of genes encoded in 31 kb repeat region is greater in roots from SCN-resistant compared to SCN-susceptible varieties. Mean ± std. error of mean shown for qPCR; results for *Glyma18g02600* were at limit of detection.

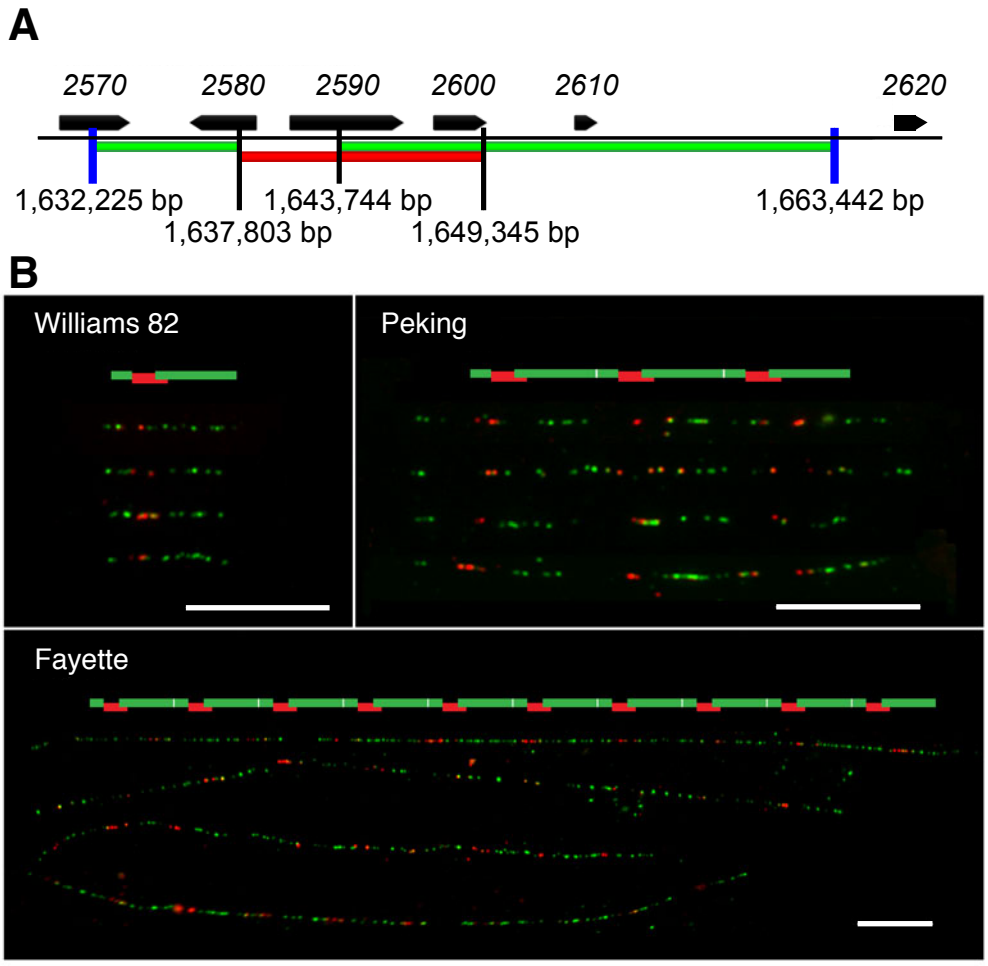


Figure. 3.

Fiber-FISH detection of *Rhg1* copy number variation in widely used soybean lines. **(A)** Two adjacent probes were isolated from a single PI88788 (*rhg1-b*) genomic DNA fosmid clone whose insert spans a repeat junction, generating the 25.2 kb (green label) and 9.7 kb (red label) probes that correspond to the Williams 82 chromosome 18 sequence as shown. **(B)** Probe diagram and composite of four Fiber-FISH images (four DNA fibers) per genotype, revealing ten or three direct repeat copies of the 31 kb *Rhg1* segment in SCN-resistant Fayette and Peking and one copy per *Rhg1* haplotype in SCN-susceptible Williams 82. White bars = 10 µm, which correspond to approximately 32 kb using a 3.21 kb/µm conversion rate (Cheng et al., 2002).

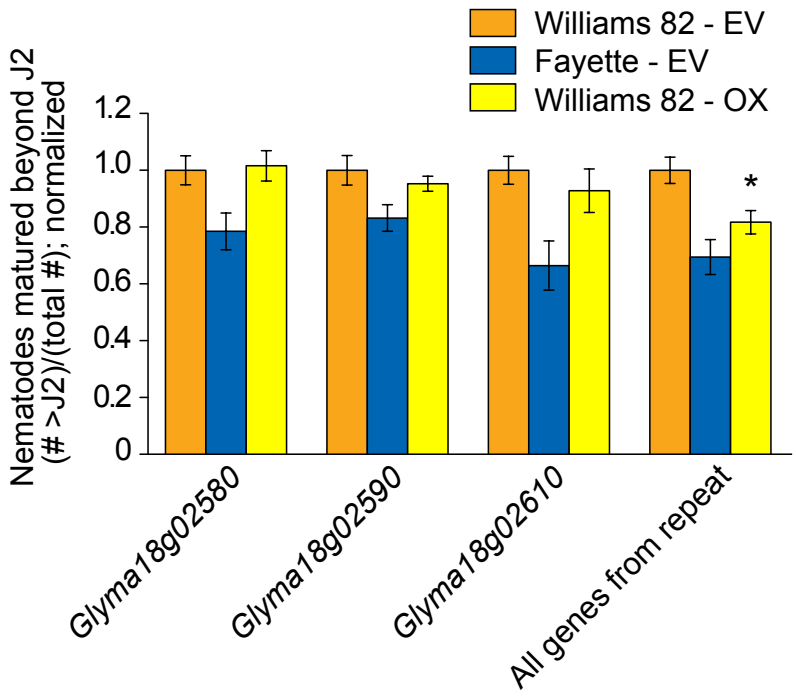


Figure. 4.

Elevated SCN resistance conferred by simultaneous overexpression of multiple genes rather than overexpression of individual genes from the 31 kb *rhg1-b* repeat. SCN development beyond J2 stage is reported for transgenic soybean roots (variety Williams 82) overexpressing the designated single genes, or overexpressing all genes encoded within the 31 kb repeat (*Glyma18g02580*, *-2590*, *-2600* and *-2610*), relative to Williams 82 (SCN-susceptible) and Fayette (SCN-resistant) controls. Mean ± std. error of mean for roots transformed with empty vector (EV) or gene overexpression constructs (OX). *: Statistical significance measured using ANOVA and Tukey test $p < 0.05$.

3.6. Supplemental Figures

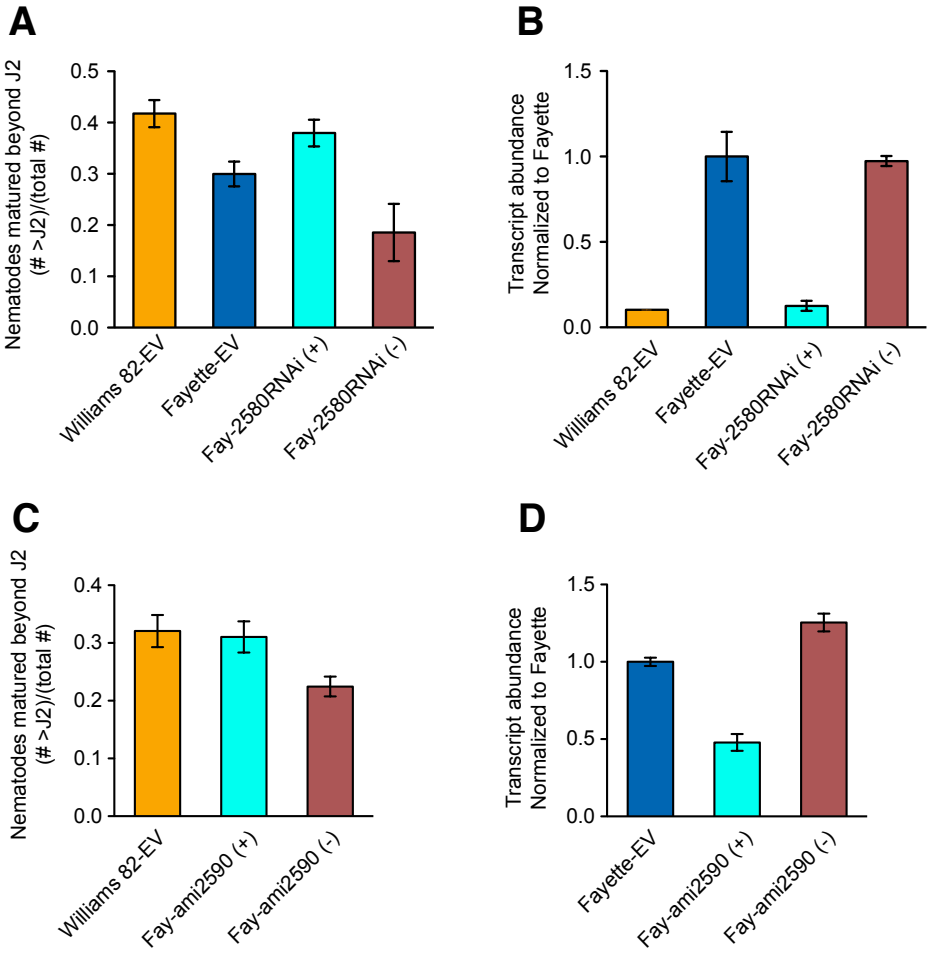


Figure S1.

Nematode development is impacted by level of silencing. **(A, C)** Nematode development on Williams 82 and Fayette roots transformed with empty vector (EV), or Fayette transformed with silencing constructs (2580RNAi or ami2590) was dependent on level of silencing. Transgenic roots with reduced target transcript abundance (+) displayed nematode development similar to Williams 82 (SCN-susceptible), while transgenic roots with non-silenced transcript level (-) had nematode development similar to Fayette (SCN-resistant). **(B, D)** Transcript abundance of target genes in roots from (A) or (C) respectively, measured by qPCR. *SKP16* transcript used as reference and normalized to Fayette-EV. The results of (B) and (D) were used to place roots in the 'well-silenced' (+) or 'not well-silenced' (-) categories shown in (A) and (C). **(A, B)** *Glyma18g02580.* **(C, D)** *Glyma18g02590.* Bars represent mean ± std. error of mean.

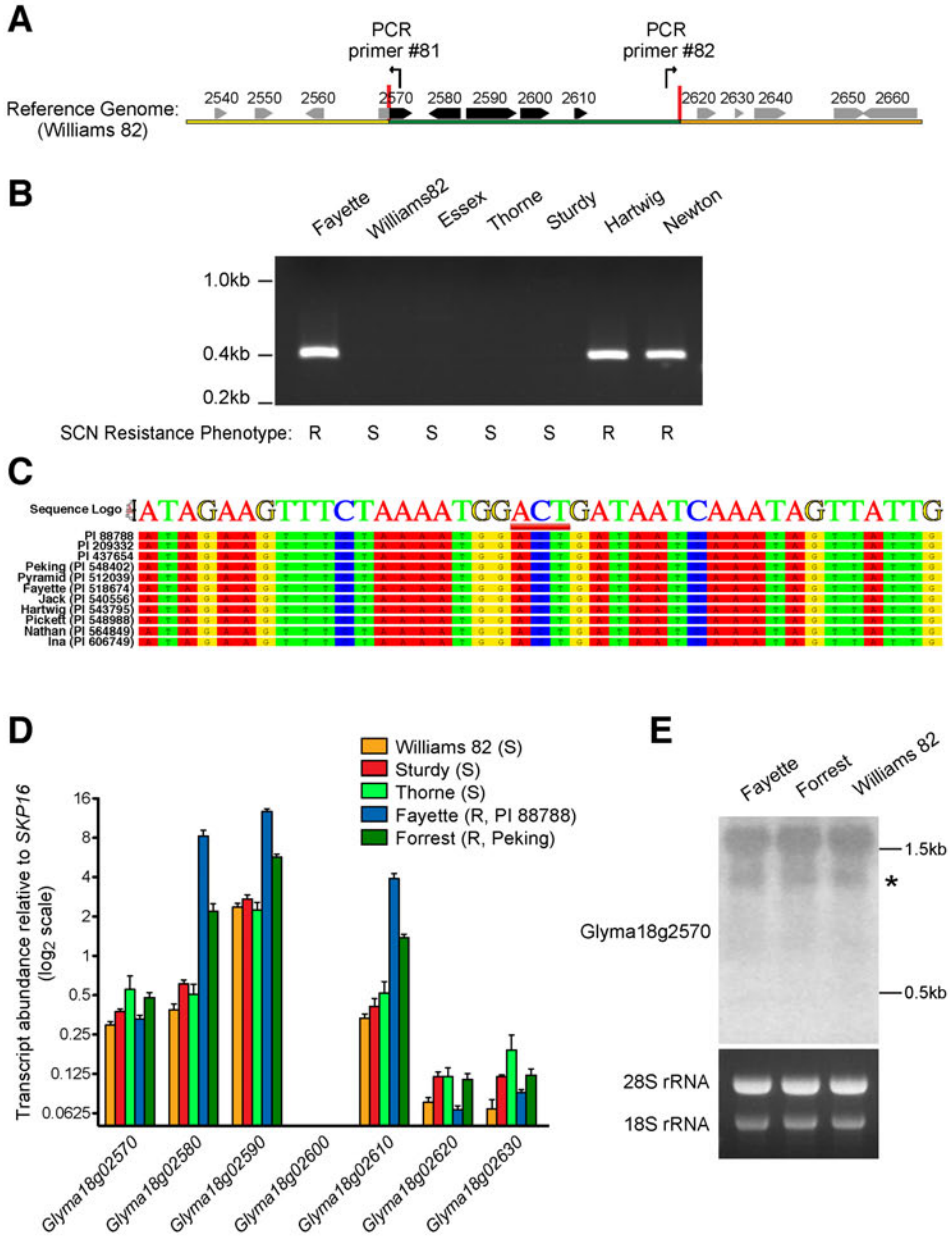


Figure S2.

Multiple SCN-resistant varieties contain the DNA junction indicative of a repeat within the *Rhg1* locus, and exhibit elevated expression of genes fully encoded within the repeat. **(A)** Schematic of PCR primers used in (B) (see also Figure 2). **(B)** Results of PCR using

Figure S2 cont'd
outward-directed oligonucleotide primers shown in (A) that match sequences at the outer edges of the 31 kb segment of *Rhg1* locus that is repeated in some soybean varieties. R indicates SCN-resistant and S indicates SCN-susceptible soybean variety. For primers 81 and 82 see Table S3. **(C)** DNA Sequence from 11 SCN-resistant varieties reveals identical sequence for repeat junction indicating a shared origin. Red bar indicates repeat junction (see also Figure 2). The single size of all junction-amplification PCR products and the consistency of all junction sequences assembled from fosmid or genomic DNA sequencing further suggest the presence of adjacent direct repeat copies. **(D)** Transcript abundance for genes encoded at *Rhg1* (normalized to *SKP16*), revealing elevated expression of genes fully encoded within the repeats of *Rhg1* from PI 88788 or Peking sources, relative to expression of the same genes in SCN-susceptible varieties. Bars represent mean ± std. error of mean. *Glyma18g02600* is expressed below 0.01% of *SKP16* (CT > 35 cycles). **(E)** RNA blot analysis for *Glyma18g02570* using RNA collected from roots of whole plants of Fayette and Forrest (SCN resistant) and Williams 82 (SCN susceptible). * denotes the band corresponding to the expected transcript size of *Glyma18g02570* (1.2 kb). The band at 1.8kb corresponds to non-specific ribosomal binding. Cultivars Fayette and Forrest (that contain repeats of the 31kb DNA segment) display the same banding pattern as Williams 82 (that contains a single copy of the 31kb DNA segment); no alternative transcripts for *Glyma18g02570* were detected as a result of the repeated DNA in Fayette and Forrest. RACE PCR from plants carrying *rhg1-b* confirmed full-length transcripts (with transcript ends as annotated in the reference genome) for *Glyma18g02580*, *-2590* and *-2610* (M.M., 2012)

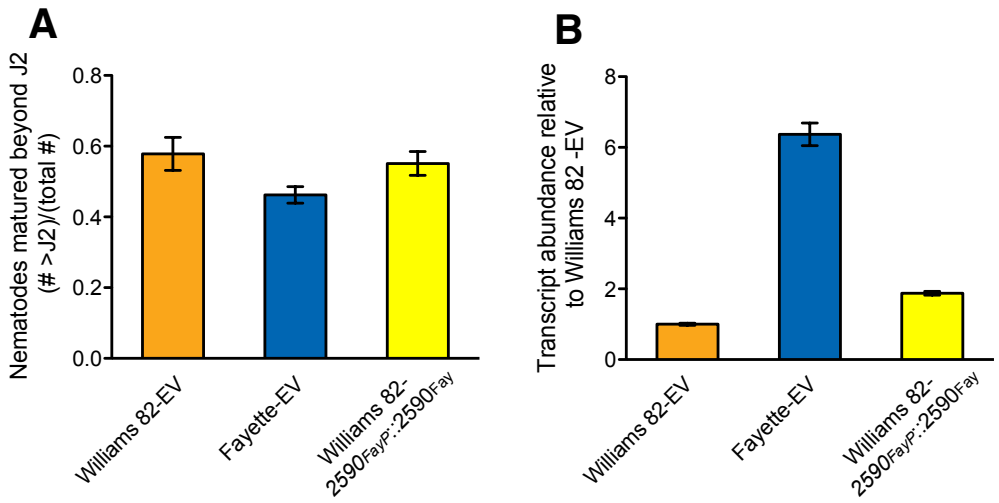


Figure S3.

Expressing the native Fayette *Glyma18g02590* allele in Williams 82 does not alter SCN development. **(A)** Similar nematode development on transgenic roots of Williams 82 expressing empty vector (EV) or Williams 82 expressing the Fayette (*rhg1-b*-type) allele of *Glyma18g02590* under control of Fayette *Glyma18g02590* promoter sequences ($2590_{FayP}::2590_{Fay}$). Williams 82 transformed with either construct allowed a greater proportion of nematodes to advance beyond the J2 stage compared to Fayette-EV. **(B)** Transcript abundance for *Glyma18g02590* in roots from (A), measured by qPCR. *SKP16* transcript used as reference; data normalized to Williams 82 - EV. Bars in (A) and (B) represent mean ± std. error of mean.

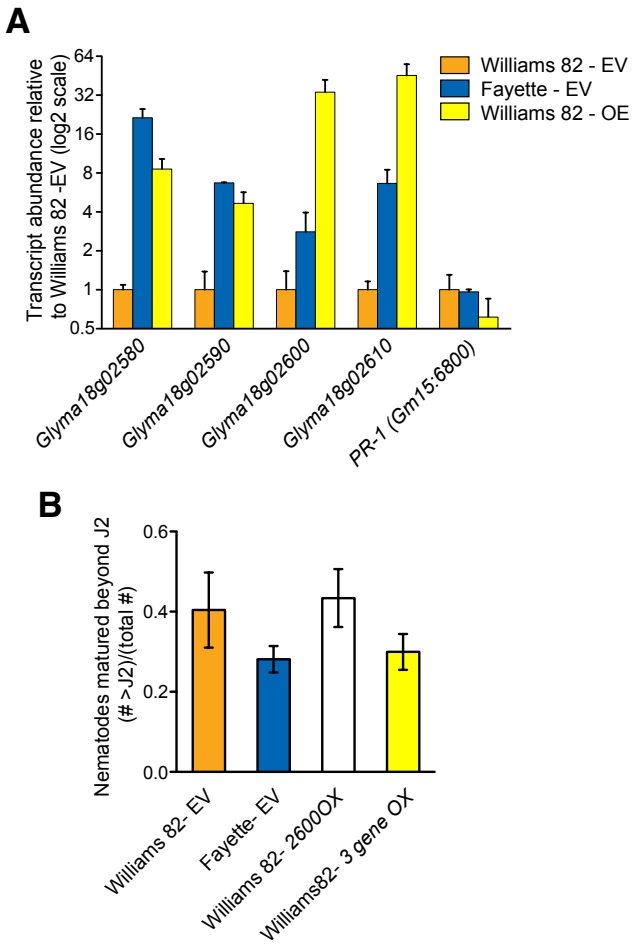


Figure S4.

(**A)** qPCR reveals elevated transcript abundance of the intended genes in roots transformed with the multiple gene simultaneous overexpression construct of Figure 4, and no significant elevation of PR-1 expression that might otherwise indicate non-specific elevation of defenses. Transgenic roots carried either the multiple-gene construct (OX) or empty vector (EV). Similar results obtained in second independent experiment with different transgenic events, except PR-1 abundance was more similar (closer to 1.0) between Williams 82 - EV, Fayette-EV and Williams-OX roots in second experiment. Bars represent mean ± std. error of mean. Data for *Glyma18g02600* are less dependable for Williams-EV and Fayette-EV because their qPCR signal was at the limit of accurate qPCR detection (CT > 33). **(B)** A smaller percentage of nematodes matured beyond J2 on Williams 82 roots overexpressing three-gene construct, in an experiment in which one gene (*Glyma18g02600*) or three genes (*Glyma18g02580, -2590* and *-2610*) were overexpressed, compared to SCN-resistant Fayette and SCN-susceptible Williams 82 controls. Left to right, n = 7, 4, 7, 8 transgenic roots respectively.

3.7. Supplementary Tables

Table S1.

Length estimates for Fiber-FISH hybridization signals. N: number of DNA fibers analyzed.

	Length of Fiber-FISH signals (μm)	Estimated Length of Fiber-FISH signals (kb)	N
Williams 82	8.60 ± 0.49	27.60 ± 1.58	20
Fayette	96.56 ± 7.54	309.94 ± 24.2	20
Peking	34.42 ± 2.91	110.48 ± 9.34	20

Table S2.

Amino acid polymorphisms within the 31 kb repeat at Rhg1-b The position of the protein-coding gene was predicted by comparing the fosmid clone sequences to the Glyma1 version of the soybean genome assembly. (L) or (R) indicate either the left or right side of the junction. * indicates an insertion of 1 aa (3 bp of DNA sequence) between amino acid position 287 and 288 based on the Williams 82 genome assembly (Glyma1; www.phytozome.net)

Gene ID	Amino Acid Position	W82	#1	#3	#4(L)	#5(L)	#5(R)
2590	203	Q	K	Q	K	-	K
2590	285	E	Q	E	Q	Q	Q
2590	286	D	H	D	H	H	H
2590	287	D	E	D	E	E	E
2590	287-288*	-	A	-	A	A	A
2590	288	L	I	L	I	I	I

Table S3.

DNA sequences of oligonucleotide primers used for PCR (3 pages).

ID	Primer	Sequence
		Silencing constructs
1	2570 hpRNAi _F	AGGATCCATTTAAATCAAGTACTCTTCCCCACAAAAGCT
2	2570 hpRNAi _R	ACCTAGGAGGCGCGCCTGGGGCCATTTCAGTAATTAGGTC
3	2580 hpRNAi _F	acctaggaggcgcgccTCATGAAGGTTCTCGGCGTAG
4	2580 hpRNAi _R	aggatccatttaaatCCACCAGTGAATTCCAAACCA
5	2590 hpRNAi _F	GAcctaggcgcgccGGACTTGGTCGTCAACACAGTC
6	2590 hpRNAi _R	GCggatccatttaaatGAGCAGCAAACTGGGCAACT
7	2600 hpRNAi _F	acctaggaggcgcgccGCCAAATTCAAAAGGCTTGCT
8	2600 hpRNAi _R	aggatccatttaaatCACCATTCAACATGCCTGTCA
9	2610 hpRNAi _F	taacctaggaggcgcgccACAACTCCTTCCGATTCGTTCCG
10	2610 hpRNAi _R	caggatccatttaaatAGATACAACCACCTGAATACGCCC
11	2620 hpRNAi _F	AGGATCCATTTAAATCTCGCAACACCATATCCAGAGTA
12	2620 hpRNAi _R	ACCTAGGAGGCGCGCCGGTGTTAAGGTCGAACCTGCGAA
13	2590-1 I miR-s	gaTATTGGTTATAGCAACACCGTtctctcttttgtattcc
14	2590-1 II miR-a	gaACTTTGCTATAACCAATAtcaaagagaatcaatga
15	2590-1 III miR*s	gaACAGTGTTGCTATTACCAATTtcacaggtcgtgatatg
16	2590-1 IV miR*a	gaAATTGGTAATAGCAACACTGTtctacatatatattcct
17	2610-1 I miR-s	gaTATTTCCCGACCCGACGGGACtctctcttttgtattcc
18	2610-1 II miR-a	gaGTCCCGTCGGGTCGGGAAATAtcaaagagaatcaatga
19	2610-1 III miR*s	gaGTACCGTCGGGTCCGGAAATTtcacaggtcgtgatatg
20	2610-1 IV miR*a	gaAATTTCCGGACCCGACGGTACtctacatatatattcct
21	2610-2 I miR-s	gaTATCCAGTCACCGCGACGTGGtctctcttttgtattcc
22	2610-2 II miR-a	gaCCACGTCGCGGTGACTGGATAtcaaagagaatcaatga
23	2610-2 III miR*s	gaCCCCGTCGCGGTGTCTGGATTtcacaggtcgtgatatg
24	2610-2 IV miR*a	gaAATCCAGACACCGCGACGGGGtctacatatatattcct
		Transcript Abundance using qPCR
25	gm18: 2570 F_qPCR	TGAGATGGGTGGAGCTCAAGAAC
26	gm18: 2570 R_qPCR	AGCTTCATCTGATTGTGACAGTGC
27	gm18: 2580 F_qPCR	CGTGTAGAGTCCTTGAAGTACAGC
28	gm18: 2580 R_qPCR	ACCAGAGCTGTGATAGCCAACC
29	gm18: 2590 F_qPCR	TCGCCAAATCATGGGACAAGGC
30	gm18: 2590 R_qPCR	CAATGTGCAGCATCGACATGGG
31	gm18: 2600 F_qPCR	GCTTCAGTCAAGAAAATGTGCATG
32	gm18: 2600 R_qPCR	CACCCGAAACCGCGACACAAATG
33	gm18: 2610 F_qPCR	AGGTCACGTGTTGCCGTTG
34	gm18: 2610 R_qPCR	AAACCACACCAATAACAACAAAGCTCT
35	gm18: 2620 F_qPCR	AAGCCCAACAGGCCAAAGAGAG
36	gm18: 2620 R_qPCR	ACACCAAATGGGTTCGCACTTC
37	gm18: 2630 F_qPCR	TTGTGGAAGTGAAAGTCGGTTTGC
38	gm18: 2630 R_qPCR	GTTGTCACGTTTCCCGTAACAATG

39	EF1b_For.qRT	CCACTGCTGAAGAAGATGATGATG
40	EF1b_Rev.qRT	AAGGACAGAAGACTTGCCACTC
41	SKIP16_For.qRT	GAGCCCAAGACATTGCGAGAG
42	SKIP16_Rev.qRT	CGGAAGCGGAAGAACTGAACC
43	UKN2_For.qRT	GCCTCTGGATACCTGCTCAAG
44	UKN2_Rev.qRT	ACCTCCTCCTCAAACTCCTCTG
45	ACT11_For.qPCR	ATCTTGACTGAGCGTGGTTATTCC
46	ACT11_Rev.qPCR	GCTGGTCCTGGCTGTCTCC
47	UNK1_For.qPCR	TGGTGCTGCCGCTATTTACTG
48	UNK1_Rev.qPCR	GGTGGAAGGAACTGCTAACAATC
49	TIP41_For. qPCR	AGGATGAACTCGCTGATAATGG
50	TIP41_Rev.qPCR	CAGAAACGCAACAGAAGAAACC
51	PR-1 (6790) F	TGCTTGGTCACCTGGAAGTTGG
52	PR-1 (6790) R	AACTTCCTGCGAGCTGCGATAC
53	PR-1 (6800) F	AGTCATTGTGGGTGATCATGCTG
54	PR-1 (6800) R	GCAGCGTTGTGTGCATTAACAAAG

Expression Vectors

55	Ox2610-SalF2	gtcgacATGCGCATGCTCACCGG
56	P2610fused-R	TATTGCGAGAACCAAACCGG
57	Ox2590Sal-F	GGgtcgacATGGCCGATCAGTTATCGAAGG
58	Ox2590fused-R	AGTAATAGCCTCATGCTGCTCAAGTT
59	TerXba-R	ACtctagaGCGCATGTCTTGCGTTGATG
60	GmubiXba-F	GCtctagaGGGCCCAATATAACAACGACG
61	TerKpn-R	TCggtaccGCGCATGTCTTGCGTTGATG
62	PPA Linker_Top	GatgtcTTAATTAAtatctgtGGGCCCactatGGCGCGCCaatgtaaA
63	PPA Linker_Bottom	AGCTTttacattGGCGCGCCatagtGGGCCCacagataTTAATTAAgacatCTGCA
64	Ox2600-F	ATGGTTTCGGTTGATGATGGG
65	Ox2600-R	TTTTTGTGCATATAAGGGGTTCAT
66	NosHind-F	GCaagcttGATCATGAGCGGAGAATTAAGGG
67	Nos2600-R	CCCATCATCAACCGAAACCATAGATCCGGTGCAGATTATTTGG
68	Nos2600-F	CCAAATAATCTGCACCGGATCTATGGTTTCGGTTGATGATGGG
69	NosAsc-R	TCggcgcgccGCGCATGTCTTGCGTTGATG
70	Ox2580-F	ATGTCTCCGGCCGCCG
71	Ox2580-R	TGACTTGCTACTAAAAGCATTATATATGTTG
72	NosAsc-F	CAggcgcgccGATCATGAGCGGAGAATTAAGGG
73	Nos2580-R	CGGCGGCCGGAGACATAGATCCGGTGCAGATTATTTGG
74	Nos2580-F	CCAAATAATCTGCACCGGATCTATGTCTCCGGCCGCCG
75	NosSbf-R	TGcctgcaggGCGCATGTCTTGCGTTGATG
76	M13 F	GTAAAACGACGGCCAG
77	M13 R	CAGGAAACAGCTATGAC
78	pSM101 seq	GTCTTGATGAGACCTGCTGCG
79	g2590pHind-F	CTTaagcttGAATGGTTTTTGTTTTGTTGTCTCTCAC
80	g2590pSal-R	TTGGTCGACCGTATCATCCAATG

Bridge PCR

81	SCN_Res Bridge F	TTTAGCCTGCTCCTCACAAATTC

82	SCN_Res Bridge R	TTGGAGAATATGCTCTCGGTTGT
	Probes for Northern Analysis and Alt Transcript	
83	2570F_qPCR	TGAGATGGGTGGAGCTCAAGAAC
84	2570 UTR Rev	CAAGTACTCTTCCCCACAAAAGC
85	2570 put exon F	TGCAGTTTTAGTGGAAAGGCC
86	2570 exon 6 R	TCATCAAGCTCAACTTGAATCCC
	RACE PCR	
87	2590-5GSP	GATCGGCCATTTTCCTCCGATCGAAACA
88	2590-5NGSP	GACGACCAAGTCCAAATCCAAAACCCGC
89	2590-3GSP	AAGCCAAAGAACTTGAGCAGCATGAGGC
90	2590-3NGSP	CTGTCCAGTTGTTCGTCTTACACATCCA
91	2610-5GSP	GGCGACGATCTTGACGACGGCGTT
92	2610-5NGSP	TCATACAGTGCAACCACCAGCCGCG
93	2610-3GSP	GGACGAGGTCACGTGTTGCCGTTGCT
94	2610-3NGSP	TTCACCACTATGGGCGTATTCAGGTGGT
95	2580-3GSP	CCTGGGGGATTCCAAAGGAACGC

4. Population analysis of *Rhg1* identifies complex polymorphisms for copy number, coding sequence, and DNA methylation states directly contribute to soybean cyst nematode resistance

Data for this chapter are being prepared for submission.

I conducted qPCR estimates for transcript abundance, collected and isolated DNA for genomic sequencing with Adam Bayless, cloned soybean transcripts along with Adam Bayless, analyzed the whole genome sequence data, performed the methylation survey in soybean using restriction enzymes with Xiaoli Guo, and analyzed the data from the soybean methylome. Adam Bayless conducted the initial copy number estimates using qPCR and isolated the truncated 2590 allele, along with previously mentioned contributions. I helped design many of the experiments, wrote the manuscript and produced the figures (except those detailing fiber-Fish, performed by Kia Wang) in conjunction with Andrew Bent.

4.1. Abstract

Ongoing adaptation and evolution are critical in order for organisms to thrive in the face of local environmental variables. Plant breeders have long taken advantage of this process to develop better crops by identifying natural sources of tolerance to biotic and abiotic stresses. Here we describe the genetic variation for one such trait in soybean – *Rhg1*-mediated resistance to soybean cyst nematode, the most economically damaging pathogen of soybean. Using transcript profiling, fiber-FISH, and whole genome sequencing in diverse population of over forty soybean lines, we have identified one susceptible genotypic class and two distinct resistance classes for soybean cyst nematode resistance encoded at the *Rhg1* locus. Copy number variation, transcript abundance, nucleic acid polymorphisms, and differentially methylated DNA regions define the two classes that represent the majority of the available soybean cyst nematode resistance in commercially cultivated varieties. All tested lines belonging to the high-copy class of *Rhg1*, including PI 88788, are copy number variants that contain between 7 to 10 copies of the 31 kb *Rhg1* repeat, and contain the previously identified alternate allele of *Glyma18g02590*, a predicted α-SNAP. We have also defined a set of low-copy class lines, including PI 548402 or Peking, which all contain 3 tandem repeats of the *Rhg1* locus, and a new discovered alternative allele of the *Glyma18g02590* α-SNAP. There is strong evidence for a shared origin of the *Rhg1* repeats, but subsequent bifurcated evolution of the high-copy and low-copy *Rhg1* loci. Differentially methylated DNA regions were identified that correlate with soybean cyst nematode resistance and likely play a role in the expression and evolution of the locus. Together, these data represent an

extensive characterization of an agronomically important adaptive trait, and provide evidence for evolution of stress tolerance by gene copy number variation.

4.2. Introduction

Following the origin of terrestrial land plants, vascular plants have experienced rapid diversification and now occupy a diverse range of environments, overcoming a number of abiotic and biotic stresses (Kenrick and Crane, 1997; Steemans et al., 2009; Oh et al., 2012). The sessile life history of land plants also creates strong selection pressure to evolve and survive local environmental challenges. One mechanism by which genetic variation can be created is through the diversification of the physical genome, broadly referred to as structural variation (Feuk et al., 2006), which is increasingly recognized for having significant impacts on phenotypes and evolution (Aitman et al., 2006; Perry et al., 2008; Maron et al., 2013). A recent study in *Zea maize* identified an association between structural variation and agronomically important traits, suggesting beneficial phenotypic diversity has been selected for in modern cultivated lines (Chia et al., 2012).

A specific type of structural variation, termed copy number variation, is particularly interesting for its varied evolutionary impact, including neofunctionalization, subfunctionalization, and gene dosage effects (Ohno, 1970; Moore and Purugganan, 2005; Flagel and Wendel, 2009; Marques-Bonet et al., 2009). While the majority of duplicated genes do not retain a function due to a process of pseudogenation (Lynch and Conery, 2000; Demuth and Hahn, 2009), evolution through gene duplication has impacted the genomes of diverse organisms (Kondrashov et al., 2002) reviewed in (Conant and Wolfe, 2008). A number of examples detailing evolution through gene duplication as an adaptation to local environmental conditions have been reported from a

range of organisms (Triglia et al., 1991; Labbe et al., 2007; Dassanayake et al., 2011), reviewed in (Kondrashov, 2012).

Domesticated soybean, *Glycine max*, is an important world commodity, accounting for a majority of the world's protein-meal and oilseed production (soystats.com), with a reported value of soybean produced in the United States exceeding $35 billion in 2011. This importance has led to intense selection for higher yielding, more efficient, and more resistant varieties capable of overcoming biotic and abiotic stress. One such trait that has received significant attention is resistance to the soybean cyst nematode, *Heterodera glycines* (Concibido et al., 2004; Brucker et al., 2005; Kim et al., 2010; Kim et al., 2011). Soybean cyst nematode (SCN) is the number one economically damaging pathogen of soybean throughout the United States (Wrather and Koenning, 2009). Soybean cyst nematodes are obligate endoparasites that cause disease by reprogramming host root cells to form specialized feeding cells, termed syncytium, robbing the plant of carbon and adversely affecting yield (Lauritis et al., 1983, Endo, 1984 #2697; Young, 1996; Sharma, 1998). Soybean cyst nematode is found in all major soybean-growing states in the US and cannot feasibly be removed (Niblack, 2005). The number one control strategy is identification and development of soybean germplasm with natural host resistance. Two loci, *Rhg1* and *Rhg4* (Resistance to *Heterodera glycines*) make the greatest contribution to resistance in the vast majority of modern soybean cultivars that exhibit SCN resistance (Calwell et al., 1960; Matson and Williams, 1965; Webb et al., 1995; Li et al., 2004; Brucker et al., 2005)}.

We recently discovered that the *Rhg1* locus for SCN resistance, widely used in commercial soybean breeding, is mediated via a copy number polymorphism of a 31 kb segment of DNA containing four open reading frames (Cook et al., 2012). The most widely utilized version of the locus, derived from plant introduction (PI) 88788 and termed *rhg1-b*, was shown to contain 10 tandem copies of the locus. Only a single copy of this 31 kb block is present in the tested SCN-susceptible varieties, including the fully sequenced soybean reference genome of Williams 82. Equally surprising, three of the four genes within the repeat make contributions to SCN resistance. These genes are *Glyma18g02580* (encoding a predicted amino acid transporter), *Glyma18g02590* (encoding a predicted α-SNAP vesicle trafficking protein), and *Glyma18g02610* (encoding a protein lacking a predictable function). Two of these genes encode predicted protein sequences that are invariant between SCN-resistant and SCN-susceptible alleles, and further evidence was obtained suggesting that the contribution of those genes to resistance is attributable to expression polymorphism attributed to the described copy number variation (CNV). For the third gene (*Glyma18g02590*), which exhibits increased expression to a lesser extent, may further contribute to SCN resistance based on the presence of an alternate coding allele for the predicted α-SNAP. The other primary source of *Rhg1*-mediated SCN resistance in commercially cultivated soybean varieties is PI 548402 (commonly and throughout this paper referred to as 'Peking'). We found that the Peking *Rhg1* contains three copies of the 31 kb region, but nucleotide sequences of the genes in Peking *Rhg1* were not determined.

A well-documented epistasis occurs in Peking-derived SCN resistance, in which Peking *Rhg1* has reduced efficacy if Peking *Rhg4* is not simultaneously present (Brucker et al., 2005). The responsible gene at *Rhg4* was discovered and reported in 2012 to encode a serine hydroxymethyltransferase (SHMT) (Liu et al., 2012). Peking and PI 437654 (or Hartwig) contain an *Rhg4* allele whose product exhibits altered enzyme kinetcs at intermediate substrate concentrations. Impacts of *Rhg4* on SCN resistance are difficult to detect when resistance-associated *Rhg4* alleles are deployed together with the ten-copy *rhg1-b* from PI 88788 (Brucker et al., 2005). It is intriguing that SCN populations can arise that partially overcome the resistance mediated by one source of *Rhg1* yet are still successfully resisted by other *Rhg1* sources (Niblack et al., 2002; Colgrove and Niblack, 2008). In addition to the interest in understanding the biology of trait variation caused by copy number variation, and in traits conferred by tightly linked blocks of distinct genes, there is substantial practical interest in understanding the variations in SCN resistance caused by different sources of *Rhg1*, and in the potential to predict and then discover or develop more effective versions of *Rhg1*.

Here we use qPCR, fiber-FISH, whole-genome sequencing and bioinformatics analyses to investigate the major SCN resistance locus *Rhg1* from a diverse a population of soybean lines. We sequenced the genomes of six “Hg-typing” soybean lines widely used as sources of SCN resistance in breeding (Niblack et al., 2002), along with analyzing whole genome sequence from the 41 soybean SoyNAM lines to characterize nucleic acid polymorphisms (Cregan and Diers, unpublished). We discovered three classes of the *Rhg1* locus that can be differentiated by gene expression level, copy

number, and coding sequence. We explore the evolutionary history of the locus and provide data showing copy number influences this adaptive trait, and also show that differentially methylated DNA regions at *Rhg1* correlate with SCN resistance. Together these data provide a detailed analysis of one of the few confirmed plant genome duplications known to contribute to an adaptive trait.

4.3. Results

4.3.1. Commonly used sources for *Rhg1* resistance possess either a low-copy or high-copy number of *Rhg1* repeats as compared to the wild-type single copy

We recently found that the *Rhg1* loci derived from PI 88788 and Peking contain ten and three copies respectively of a 31 kb DNA segment encoding three genes that contribute to SCN resistance (Cook et al., 2012). To assess the natural variation present at *Rhg1* in commonly used soybean germplasm, we analyzed five other SCN lines that make up the established SCN type test (refered to as the Hg-type test) (Niblack et al., 2002). Initial characterization of *Rhg1* copy number, using qPCR on genomic DNA, revealed three copy number classes: single-copy, low-copy (two to six copies) and high-copy (>6 copies) (Figure 1A). For lines estimated to contain >6 copies, qPCR produced variable results and un-reliable absolute copy number estimates, possibly because it is difficult to reduce qPCR variation below ~50% (half of one PCR cycle) between replicate tissue samples. Copy number estimates based on qPCR, did however consistently identify two different classes for *Rhg1* repeats.

To determine the impact that varying *Rhg1* copy number has on transcription, we quantified root transcript abundance in the Hg-type lines using qPCR. The four genes encoded within the previously identified *Rhg1* repeat, *Glyma18g02580, Glyma18g02590, Glyma18g02600,* and *Glyma18g02610* are more highly expressed in each of the seven tested Hg-type SCN resistance lines, relative to SCN-susceptible Williams 82 (Figure 1B). The transcript abundance of an adjacent gene that is outside of the 31 kb repeat,

Glyma18g02570, was similar across all of the tested genotypes. Four of the genotypes, Peking, PI 90763, PI 89772, and PI 437654, showed similar levels of elevated expression of the repeated genes, while expression was most elevated and similar in Cloud (PI 548316), PI 88788, and PI 209332 (Figure 1B). These groupings were the same as those identified for qPCR estimates of DNA copy number, and indicate that transcript abundance for these genes is scalable to gene copy number. One gene in the repeat, *Glyma18g02600,* showed a slightly different expression pattern, in that it has elevated transcription in the resistant lines but was similarly expressed between genotypes in different copy number classes. That is, expression for this gene was similar in lines with 3 copies and 10 copies. Transcript abundance for this gene was close to the limit of detection for qPCR, was detected only at very low levels in published RNAseq experiments (Severin et al., 2010), and no contribution of this gene to SCN resistance has yet been demonstrated (Cook et al., 2012). Also of note, soybean line Cloud, which we show belongs to the high-copy number class, but was estimated to have fewer copies than PI 88788 and PI 209332, showed lower transcript abundance than the other two lines in the high-copy class (Figure 1B).

4.3.2. Copy number at the *Rhg1* locus in the high-copy lines is dynamic

Variation in *Rhg1* copy number and in the sequence and expression of *Rhg1* genes was further investigated. To definitively determine *Rhg1* copy number in the Hg-type lines, we performed fiber-FISH using a diagnostic pair of DNA probes that span the repeat junction and partially overlap (Cook et al., 2012; Walling and Jiang, 2012). The

representative fiber-FISH images for soybean lines PI 90763, PI 89772, and PI 437654 (Figure 2B, top 3 panels) summarize the finding that all three lines contain three copies of the 31 kb *Rhg1* locus per haplotype, arranged as a head-to-tail segmental duplication. These results confirm the copy number estimates from qPCR. Less expectedly, fiber-FISH results for soybean lines in the high-copy *Rhg1* class identified the presence of seven copies in Cloud, nine copies in PI 88788 and ten copies in PI 209332 (Figure 2B, bottom 3 panels). We had previously found that Fayette, a soybean variety containing an *Rhg1* locus originally from PI 88788, carries ten copies of the *Rhg1* repeat (Cook et al., 2012). Hence the number of *Rhg1* repeats can vary not only between haplotype classes, but also within the high-copy class. The observed number of *Rhg1* repeats is consistent with transcript estimates of the three genes shown to mediate *Rhg1* resistance, and indicates that increasing copy number increases transcription abundance for these genes. This trend is particularly clear in the transcript abundance for *Glyma18g02580* and *Glyma18g02590*. Together, these results show that a structural polymorphism resulting in either a low-copy number or high-copy number of *Rhg1* repeats in part defines the major sources of SCN-resistant germplasm.

To further discover the nature of the diversity within soybean *Rhg1*, we performed whole genome re-sequencing for the remaining six Hg-type soybean lines: Peking, PI 90763, PI 89772, PI 437654, PI 209332 and Cloud (Niblack et al., 2002). Derivatives of PI 88788 had previously been sequenced (Cook et al., 2012). In addition, we analyzed whole genome Illumina data that has been generated as part of the ongoing SoyNAM project (Cregan and Diers, unpublished), and previously published Illumina

sequencing data from the undomesticated, *Glycine soja* (Kim et al.). For the Hg-type line sequencing, each genome was re-sequenced (101 bp reads) using both a paired-end and a mate-pair library, generating from 9,389 Mb to 47,426 Mb of sequence per genome (Supplemental Table 1). The SoyNAM lines used for this research project were paired-end sequenced (150 bp) with sequence ranging from 5,344 Mb to 14,566 Mb per genome (Supplemental Table 1). A total of 25,464 Mb of Illumina paired-end reads from 77bp and 36 bp sequencing reactions was analyzed from *Glycine soja* (Kim et al.). For the present study we focused on in-depth analysis of *Rhg1* on chromosome 18 and its paralogous locus on chromosome 11.

To determine the prevalence of architectural variation present at *Rhg1*, we rapidly screened the SoyNAM genome sequence data sets by aligning Illumina reads to a portion of the Williams 82 reference genome corresponding to *Rhg1* and the related sequences on chromosome 11, 5, and 8 (See Methods). This screen determined that 8 of 36 SoyNAM lines contain an estimated *Rhg1* copy number greater than 1, based on read depth across the known repeat and flanking regions (Supplemental Table 2). Five SoyNAM lines previously identified as susceptibility to SCN were not mapped. As a control, copy number of the paralog on chromosome 11 was also estimated, and did not differ from one.

To further investigate the extent of copy number variation in the population of soybean genomes and to eliminate possible mapping bias that might arise from use of a limited reference sequence region, Illumina sequencing reads for all the samples were re-mapped to the entire reference genome, allowing for *Rhg1* copy number estimates based

on read depth. Along with the six Hg-type lines sequenced as part of this work, and previously sequenced *G.soja*, 24 of the 41 SoyNAM lines were selected for further analysis based on the results of the rapid alignment and on sequencing coverage. The estimated copy numbers based on read depth for the six Hg-type lines are in agreement with the results from qPCR estimates and fiber-FISH. Lines Peking, PI 90763, PI 89772, and PI 437654 contain 3 copies, while Cloud has 7 and PI 209332 has 10 (Figure 2C). Eight of the SoyNAM lines contain a copy number greater than one for the *Rhg1* locus, consistent with the estimates from the rapid mapping. Seven lines contain 9 to 10 copies of the locus, while one line contains an estimated 3 copies. These results are in agreement with pedigree information where it is publicly available. Four of the seven lines estimated to have 9 to 10 copies of *Rhg1* have pedigree's tracing back to PI 88788, and while full pedigree information is not available for the other lines, they have all been identified as SCN resistant and share resistant profiles similar to PI 88788 and PI 290332 (Cary and Diers, 2012). The line with 3 copies has both Peking and PI 437654 in its pedigree.

Data from these three independent approaches detail the copy number variation present at *Rhg1* in a diverse population of soybean lines. These results indicate that the increased copy number at *Rhg1* is not a common phenomenon in the *Glycine* genome and can be traced to a limited number of parental lines. There is also no indication that structural variation has occurred at the paralogous locus on chromosome 11 (Figure 2D). All of the Hg-type soybean lines are by definition resistant to SCN, but it is important to note that all eight of the SoyNAM lines containing multiple copies of *Rhg1* are resistant to SCN, while lines not containing multiple copies are susceptible based on data

associated with germplasm information (GRIN) and results from multiple soybean regional trials from (Cary and Diers, 2010; Cary and Diers, 2011, 2012), further supporting that the presence of an *Rhg1* repeat is indicative of SCN resistance.

4.3.3. Extensive *Rhg1* sequence variation is rare in protein coding regions but significant amino acid polymorphisms are present in the predicted α-SNAP

The whole genome sequence data were also analyzed for nucleic acid and derived amino acid variation. Genomic DNA sequence variations, including single nucleotide polymorphisms (SNPs), insertions, and deletions, were identified using the genome analysis tool-kit (GATK) pipeline (McKenna et al., 2010; DePristo et al., 2011). A total of 409 DNA variant sites across the 31.2 kb (1632223 – 1663500) *Rhg1* repeat interval were identified in at least one of the 31 genomes. The average number of variants per genome for the low-copy *Rhg1* lines is 251 ± 40, the high-copy lines 260 ± 10, and the average number of DNA variants for all lines estimated to contain a single copy of *Rhg1* is 23 ± 29 (Table 1 for full results). Despite the high number of sequence polymorphisms found in SCN resistant lines, few of the polymorphisms affected protein-coding sequences- no polymorphisms were detected in the derived amino acid sequences of *Glyma18g02580* or *Glyma18g02610*, across all SCN resistant lines (Table 2). Strikingly, two resistant allele types with distinct mutations that occur at similar sites were observed in the derived amino acid sequences of *Glyma18g02590*, one associated with the low-copy class of *Rhg1* haplotypes and the other previously reported sequence from Fayette, was associated with the lines in the high-copy class (Figure 3). The gene *Glyma18g02590*

encodes a predicted α-SNAP, which has the canonical function of stimulating NSF ATPase activity to assist in the disassembly of SNARE components following vesicle-mediated transport (Morgan et al., 1994; Barnard et al., 1997; Rice and Brunger, 1999). The amino acid sequence alignment based on the analyzed genome sequences shows the invariant structure of α-SNAPs across soybean lines, except for the alleles present in the high and low classes of SCN resistant lines. The accumulation of amino acid substitutions at similar residues between the low and high-copy class of repeats suggests a shared origin, and strongly suggest a functional importance of these sites for SCN disease resistance. The accumulation of amino acid differences at the C-terminal residues is unexpected given the finding that these residues are essential for normal SNARE disassembly in heterologous experiments (Barnard et al., 1996).

We performed 3'RACE from at least 7 independent clones for each of the Hg-type lines and Williams 82 to confirm expression of the variant *Glyma18g02590* alleles. The novel (non-Williams 82) *Glyma18g02590* alleles were identified from cDNA of the respective lines corresponding to the low or high-copy class (Supplementary Table 3). Interestingly, a low proportion of cDNA clones sequenced from PI 88788 and Cloud in the high-copy class were the wild-type (Williams 82-type) α-SNAP sequence, consistent with the identification of a low frequency of heterozygous variants being Williams sequence in these genomes (see below). Also, we did not detect wild-type α-SNAP sequence from any cDNAs in the low-copy class lines, consistent with the genomic DNA sequence data. However, a splice-isoform was consistently identified in all of the low-

copy class genomes, not found in the high-copy class genomes (Supplementary Table 3, Supplementary Figure 1).

A naturally occurring truncated allele of α-SNAP was recently identified and implicated in SCN disease resistance conferred by Peking and PI 437654 type resistance, but not PI 88788 (Matsye et al., 2012). Our results from whole genome sequencing indicate however, that the sequence encoding that truncated α-SNAP is not encoded by the gene at *Rhg1* on chromosome 18, but rather by *Glyma11g35820* - its paralog on chromosome 11 (Supplementary Figure 1). The SNPs at *Glyma11g35820* responsible for encoding the truncated allele were also identified in the high-copy class *Rhg1* SCN resistant lines Cloud and LG05-4292, which is from the SoyNAM project. Another variant of note was the identification in the Peking genome of a nucleotide deletion in the 2nd exon of *Glyma18g02600* coding sequence, observed as a heterozygous deletion (see below). Translation of the resulting mRNA results in a stop codon after 8 amino acids, truncating the protein by 314 amino acids or 58% of the wild-type protein.

4.3.4. SCN resistant lines developed from shared sources indicate differing evolutionary selection over the history of the locus

To further explore the evolutionary history of the *Rhg1* locus, phylogenetic analysis was conducted using maximum parsimony analysis. Genomic DNA of the coding sequence for genes *Glyma18g02580, Glyma18g02590, Glyma18g02600* and *Glyma18g02610* including 1kb of upstream promoter were used for the analysis, and the paralogous sequences from chromosome 11 in Williams 82 were used as an outgroup.

The consensus parsimony tree from 10 possible trees indicates a clear split between the SCN resistant and SCN susceptible lines (Figure 4). The analysis further indicates distinction between the low and high *Rhg1* copy lines with high confidence as indicated by the bootstrap confidence reported at the node (Figure 4).

The related origin of the two classes of *Rhg1* resistance can also be noted intuitively, by the 147 DNA variant sites common to all of the sequenced Hg-type lines, not detected in the single copy *Rhg1* SCN-susceptible lines (Figure 5). That is, approximately 75% of the 197 variant sites, detected in at least one of the seven Hg-type lines and not present in any susceptible genome within the *Rhg1* repeat, are present in all Hg-type lines. Furthermore, each of the 147 DNA variant sites is detected in all (for low-copy class) or nearly all (for high-copy class) of the *Rhg1* repeat copies reported by the average variant allele frequency of 86% for all sites across the locus (Figure 6A, Supplemental Table 4). A small number of DNA variants were not present in any SCN susceptible genome and common within but not between the two *Rhg1* copy number classes: 10 for low-copy and 7 for high-copy (Figure 5B, 5C). There are also a small number of DNA variants private to a single line: Peking (6), PI 88788 (0), PI 90763 (1), PI 437654 (0), PI 209332 (5), PI 89772 (0), and Cloud (1).

Along with the presence or absence of particular DNA variants as noted above, the ratio of reference to alternate sequence detected at any given position varied across genomes. Analyzing the variant sequence frequency at all sites common to the high-copy *Rhg1* Hg-type lines, it is possible to gain information about the composition of the sequence content in individual copies. Based on the variant sequence frequency, the high-

copy genomes appear to contain a single copy of the Williams 82 type repeat (Figure 6A). This is seen by a high number of sites with a non-reference sequence frequency of roughly 0.9. Other sites across the genome are invariant for the alternate sequence, indicating its presence in all copies, while other sites indicate the development and propagation of variant sequence in a smaller number of the total copies. One particular DNA variant at position 1,657,025 indicates that about half of the total copies contain the alternate sequence, indicating un-equal propagation of specific copies during evolution of the locus (Figure 6A).

This is in contrast to the same analysis of the four low-copy *Rhg1* Hg-type lines, which show almost uniform alternate sequence at variant sites (Figure 6B). The low frequency of variation present at some sites is likely the result of sequencing or mapping error. The one site not following this trend at position 1,660,795 has a frequency of 0.56 across the genomes, but the individual frequencies are Peking (0.39), PI 90763 (0.66), PI 437654 (0.76), and PI 89772 (0.42), which likely indicates the presence of the variant sequence in 1, 2, 2, and 1 copies respectively of the three copies in each genome.

4.3.5. Initial analysis indicates a relationship between *Rhg1* copy number classes and measured soybean cyst nematode resistance

Previous research has detailed the differences in SCN resistance observed between Peking or PI 437652 and PI 88788 resistance measured in terms of genetics, cell biology, and nematode development, but the cause for these observations have remained elusive (Arelli and Webb, 1996; Mahalingam and Skorupska, 1996; Kim et al., 1998;

Brucker et al., 2005; Kim et al., 2010; Klink et al., 2011). To address this, we analyzed data for soybean resistance to soybean cyst nematode collected from greenhouse trials by Alison Colgrove and colleagues in Terry Niblack's lab as part of the Northern Regional Soybean trail (Cary and Diers, 2010; Cary and Diers, 2011, 2012; Cary and Diers, 2013). The analysis included data from a total of 97 field populations collected from the trials from 2009 to 2012 including SCN field populations from 8-10 northern U.S. states and southern Canadian provinces per year.

To characterize the similarities between the resistance reactions seen across Hg-type lines, we computed correlation coefficients. Peking correlations with PI 90763 and PI 89772 were significant and had the largest r^2 values (Table 3A), indicating for example that across many nematode populations, the extent of resistance of Peking plants to any given nematode population tended to be similar to that of PI 90762. PI 88788 had significant correlations with large r^2 values with PI 209332 and Cloud (Table 3A). This correlation has been previously documented (Colgrove and Niblack, 2008), but we now have a likely explanation for these two groupings being the similarity of the respective *Rhg1* copy number and α-SNAP allele, along with the previously defined *Rhg4*. Similar to previous reports, we did not detect a strong correlation between the female index values on PI 437654 and the other lines. This is increasingly interesting given the high similarity of PI 437654 *Rhg1* copy number and sequence to the other low-copy class lines reported here. PI 437654 is recognized for its particularly high levels of resistance against diverse nematode populations, and the cloning of additional SCN resistant QTL from PI 437654 should be of great interest (Wu et al., 2009).

To test the hypothesis that more copies of *Rhg1* confer stronger SCN resistance, we further analyzed the Northern Regional SCN resistance data for the high-copy lines that carry 7, 9 or 10 copies of *Rhg1* and the same allele of α-SNAP. We used as a resistance metric, the difference between the SCN resistance (reported female index) of a pair of soybean lines against any single nematode population, and then averaged these differences across the 97 field populations. Results from a one-way ANOVA indicate that Cloud (seven copies) allowed significantly more nematodes to reach the cyst stage than either PI 88788 (nine copies) or PI 209332 (ten copies) (Figure 7). Because Cloud, PI 88788 and PI 209332 lines are not isogenic at other loci, the above data are only suggestive. A controlled test of different numbers of *Rhg1* copies in a similar genetic background is needed, but our data support the general hypothesis that more copies of *Rhg1* may result in increased SCN resistance.

4.3.6. *Rhg1* loci from different sources contain differentially methylated regions that correlate with SCN resistance

In addition to determining the structural and nucleic acid variation present at the *Rhg1* locus from different sources, we investigated potential differences in DNA methylation states. In a broad survey of DNA methylation at *Rhg1*, we used DNA methylation sensitive restriction enzymes coupled with PCR to identify differentially methylated regions (DMR). The enzyme McrBC only restricts DNA at sites of methylated cytosines regardless of sequence context, while the enzyme HpaII digests DNA at the recognition sequence CCGG only if overlapping CpG methylation of the

inner cytosine does not block digestion. Hence genomic DNA digestion by McrBC followed by PCR will not produce a product if the PCR product spans methylated cytosines, while following HpaII digestion, PCR products spanning CCGG sites will only amplify if the CpG methylation blocked digestion. Using a total of 18 primer sets, we identified 7 DMRs between SCN-susceptible genomes (carrying a single-copy *Rhg1* locus) and SCN-resistant genomes (carrying high-copy or low-copy *Rhg1* loci) (Figure 8A, B, C).

More specifically, we observed significant hypermethylation in the common promoter between the outwardly transcribed *Glyma18g02580* and *Glyma18g02590* genes from Fayette containing 10 copies of *Rhg1*, while this hypermethylation was not observed in the genomes containing three copies or a single-copy of *Rhg*1 (Figure 8A). There was also a significant amount of differential hypermethylation between the tested SCN-susceptible as opposed to SCN-resistant lines upstream, downstream and within the gene body of *Glyma18g02610.* We did not observe substantial DNA methylation or DMRs in the *Glyma18g02580* or *Glyma18g02590* gene bodies in any of the tested genomes, nor did we observe substantial methylation or DMRs in any genome within or around *Glyma18g02600.* We also analyzed methylation at the *Rhg1*-adjacent but non-repeated genes *Glyma18g02570* and *Glyma18g02620* and did not observe any DMRs.

During preparation of this manuscript, a genome-wide *Glycine max* methylome study was published involving lines LDX01-1-165 (referred to here as LDX), LD00-2817P (referred to here as LD) and progeny from their cross (Schmitz et al., 2013). LD is known to have SCN resistant derived from PI 437654, while LDX contains a single copy

of *Rhg1* (Diers et al., 2010; Kim et al., 2011). To confirm our observations and gain single base methylation resolution, we highlighted and re-analyzed a portion of the data (Schmitz et al., 2013).

We re-analyzed the results of the Methyl-Seq analysis, in which sequencing results were reported as the number of times the position was sequenced along with the number of reads supporting methylation at each position, focusing on the *Rhg1* interval (See Methods). As predicted based on our findings, the *Rhg1* copy number estimate for LD was 2.93 and for LDX 1.17, with the progeny showing a range of estimates between 1 and 3 (Figure 9A). We were also able to estimate transcript abundance for the two parents along with two progeny based on the results of RNA-seq data from the study (See Methods). Again consistent with our previous findings, standardized RNA sequencing depth, normalized to the susceptible LDX parent, shows elevated expression for the genes encoded within but not adjacent to the *Rhg1* repeat in LD and progeny 11272, but not progeny 11268, consistent with *Rhg1* copy number estimates (Figure 9B).

With these results confirming the increased *Rhg1* copy number and expression, we analyzed the methylation data to confirm our findings for DMRs and to address methylation pattern heritability. A screen shot from the publicly available browser (http://neomorph.salk.edu/soybean_RIL_methylomes/browser.html) intuitively shows distinct methylation differences in the parental lines LD and LDX, specifically at the DNA after the *Glyma18g02580* coding sequence, the interval between genes *Glyma18g02580* and *Glyma18g02590*, and the regions flanking *Glyma18g02610* (Figure 10A). As predicted by the finding that methylation patterns are largely inherited based on

the parental methylation pattern (Schmitz et al., 2013), we observe similarly high levels of cytosine methylation in all sequence contexts in progeny 11272 and 11186, both predicted to have an *Rhg1* copy number greater than 2.

The authors also identified 3 CG-DMR and 5 C-DMR at the locus using an unbiased, genome wide survey, summarized in Figure 10B. The C-DMR's were also all detected as methyl-QTLs based on segregation analysis. This analysis provides base pair resolution, and statistically validated methylation differences from a genome wide survey, confirming our findings of differentially methylated cytosines between SCN resistant and susceptible genotypes.

4.4. Discussion

Soybean cyst nematode is the most economically limiting pathogen for soybean causing billions of dollars of yield losses annually in the United States alone. A major breeding effort in soybean is the continued development of better and new sources of resistance. We had previously identified a gene cluster at *Rhg1*, present in ten copies in a common SCN resistant variety along with an altered amino acid sequence for one of the four genes, contributes to SCN resistance, but the extent of *Rhg1* variation in a broader set of germplasm was unknown. Here, we report an extensive characterization of the structural, coding, and methylation differences present at *Rhg1* from a diverse population of soybean lines.

The discovery of 7, 9, and 10 copies of the *Rhg1* locus composed of nearly identical sequence across repeats and between analyzed genomes, indicates copy number at *Rhg1* is plastic, and is malleable over the time scale of breeding cycles. This is directly evidenced by the discovery of 10 copies of *Rhg1* in Fayette, a line developed by backcrossing PI 88788 (9 copies of *Rhg1*) to Williams 82 (single copy genome) (Mikel et al., 2010). This information, coupled with the analysis presented here, indicating a strong link between *Rhg1* copy numbers and increasing resistance, suggests a new strategy to increase SCN resistance through increasing the number of *Rhg1* copies could prove beneficial.

It is interesting however, that all the sequenced SCN resistant lines containing 3 copies of *Rhg1* share the same allele of α-SNAP, which is different than the allele in the

high-copy, dynamic genomes. Biochemical characterization of the two alleles is needed to determine what if any altered function they have compared to each other and wild-type function. It is interesting to speculate, however that while the genomes containing the PI 88788 type α-SNAP have experienced an increase in *Rhg1* copy number, the genomes with the Peking type α-SNAP have remained at three copies because of selection against an unknown negative impact of the Peking-type α-SNAP. In this theoretic case, the *Rhg1* locus would be under positive selection for increased copy number to improve SCN resistance, but under negative selection to maintain proper gene dosage and normal cell physiology mediated by α-SNAP. It is also possible that copy number in these genomes is affected by the observed splice isoform of α-SNAP only detected in the 3 copy *Rhg1* genomes, or the occurrence of a specific *Rhg4* allele in these genomes.

Future efforts to discover Peking type SCN resistant varieties with *Rhg1* copy numbers greater than three should help answer this questions, as will experiments to assess the effect of combining different alleles of α-SNAP into the same genome with varying *Rhg1* copy numbers. This can feasibly be carried out by cross low-copy and high-copy lines and the recovery of the rare recombinant progeny containing the two types of repeats in *cis*, or alternative biotech approaches could be used. This type of research is increasingly important given the reported erosion of the heavily leveraged PI 88788 derived resistance (Niblack et al., 2008). Continued characterization of stacked *Rhg1*copies could prove important for attenuating the development of virulent nematode populations.

Our data also provide an alternative explanation to the relatively ineffectual Peking-type *Rhg1* in the absence of the Peking-type *Rhg4*. Our data suggests that more copies of *Rhg1* can increase SCN resistance, and it is possible that the 3 copies of *Rhg1* in low copy lines is less effective at perturbing nematode development. This also suggests that *Rhg1* from Peking would provide a high level of SCN resistance on its own (without Peking-type *Rhg4*) provided copy number or expression were elevated. Previously published results in which lines containing the 3 Peking-type *Rhg1* copies are still more resistant to infection than wild-type single copy *Rhg1* lines supports this hypothesis(Brucker et al., 2005; Liu et al., 2012). Also, data in (Liu et al., 2012) shows that the 3 copy *Rhg1* locus is able to reduce nematode development by >60% in a homozygous *Rhg4* mutant line. Alternatively, when Peking-type *Rhg4* is introduced into PI 88788, there is a slight increase in resistance, but it is difficult to detect given the high level of PI 88788 resistance on its own (Brucker et al., 2005). These data suggest that the effect of *Rhg4* is masked in the resistant PI 88788 background, and that only in the sensitized Peking-type 3 *Rhg1* copy background can one observe the effect of *Rhg4* on nematode development. Collectively, these data indicate that Peking-type *Rhg1* and *Rhg4* are not necessarily required in the same genome for mechanistic function.

Determining the course of evolution at the locus is difficult, but our data suggest the repeats in the low-copy and high-copy class have a common origin based on their high sequence identity, especially in intergenic regions. One explanation for these observations is that many of the DNA variant sites within the *Rhg1* locus accumulated after an initial duplication event. Following duplication, a number of possibilities have

been described for the different outcomes of the duplicates (Conant and Wolfe, 2008). The category II model, formally described in (Innan and Kondrashov, 2010) details the evolutionary possibilities of duplicates resulting from a beneficial increase in dosage. This model could account for the large number of DNA variants found at *Rhg1*, if the original duplication had a small increase in fitness, resulting in a predicted relaxed selection of the gene copies. Following the accumulation of an increasingly beneficial variant, positive selection for increased dosage and negative selection on the coding sequences could have drove the locus to the high-copy numbers observed today (Innan and Kondrashov, 2010). Alternative models involving the complete redundancy hypothesis (Lynch and Conery, 2000), could account for the increase in accumulation of DNA variant sequences observed at *Rhg1*, followed by positive selection for increase gene dosage (Innan and Kondrashov, 2010). The observation of one copy of *Glyma18g02600* having a premature stop codon as the result of a novel upstream deletion in Peking, could be the first glimpse of psuedogenation, suggest this gene is not under the same purifying selection as the rest (Lynch and Conery, 2000). Regardless of the exact nature of selection on the locus, our data suggested increased fitness with increased copies when grown under SCN pressure.

The identification of differentially methylated DNA regions present at *Rhg1*, between SCN resistant and susceptible genomes adds an additional layer of complexity controlling the expression and possibly evolution of the locus. The observation of highly similar gene duplicates in the genomes of many organisms, has led to the hypothesis that decreased expression of duplicate gene copies is one mechanism to maintain normal

physiology following gene duplication (Qian et al., 2010). Recent work on mammalian gene duplicates suggests increased DNA methylation of promoter regions was significantly correlated with gene duplicates and silencing, suggesting a potential mechanism for the restoration of dosage imbalance (Chang and Liao, 2012). This observation was also made in soybean, in which one copy of a paralogous pair was found to have increased repressive methylation and decreased expression for many pairs (Schmitz et al., 2013). This is seemingly paradoxical to our observation of hypermethylation and increased expression of the genes encoded in the *Rhg1* repeat for SCN resistance lines with multiple copies. An explanation for these observations could be that expression of the duplicated genes would be even greater in the SCN resistant genomes without the repressive methylation, and that while expression is constitutively elevated, it is still maintained within an acceptable physiologic range. The recent identification of dynamic methylation changes in *Arabidopsis* following biotic stress (Dowen et al., 2012), further suggests that the C-DMRs found to occur upstream of *Glyma18g02580, Glyma18g02590*, and *Glyma18g02610* could result in increased expression of these genes following nematode infection. Together, these results suggest that DNA methylation has evolved at *Rhg1* repeats to maintain normally physiological gene dosage, but the ability to respond to pathogen attack by de-methylation of the several repeats could provide a potent response to pathogen infection. Future experiments to test this hypothesis could provide new evidence for the evolution of gene duplicates involving increased fitness during times of local environmental challenge.

4.5. Materials and Methods

4.5.1. Estimating copy number and transcript abundance

To estimate the number of *Rhg1* copies present in the Hg-type lines, we collected tissue for DNA extraction from two week old plants grown in metro mix at 26C. Leaf tissue was collected and flash frozen in liquid nitrogen and DNA extraction was performed as previously described. To estimate *Rhg1* copy number qPCR reactions were run with using two separate primer pairs per sample. One set of primers previously described span the junction of repeated segmental *Rhg1* duplicates, which fail to amplify a product in genomes with the wild-type single copy of the locus. A second primer pair used in a separate reaction amplified a product corresponding to a DNA interval from the gene *Glyma18g02620* which is adjacent to, but not present in the *Rhg1* repeat. The ratio of the two products was used to determine the number of *Rhg1* repeats.

To quantify the relative transcript abundance for the genes within and adjacent to the *Rhg1* repeat interval, tissue was collected from the roots of plants five days post emergence. Plants were grown in a growth room in metromix at 24 C and 16 hrs of light. The entire root, soil mass was removed from the pot, quickly immersed in water to remove excess soil and flash froozen in liquid nitrogen. RNA was extracted using Trizol following manufactures recommended procedures. Contaminating DNA was removed from the samples using Turbo DNAse following manufactures guidelines. To amplify cDNA from RNA, Biorad's iScript kit was used with 1ug of total RNA per reaction following manufactures recommended guidelines. qPCR was performed as previously

published (Cook et al., 2012). Briefly, primer pairs corresponding to transcripts of, *Glyma18g02570, Glyma18g02580, Glyma18g02590, Glyma18g02600*, and *Glyma18g02610* were used to amplify products for each sample in duplicate technical replicates. A product was also amplified from each sample corresponding to transcript of gene *SKP16* for use in normalizing samples across plates (Cook et al., 2012).

4.5.2. Transcript analysis

To confirm the annotation of transcripts at *Rhg1*, rapid amplification of cDNA ends (RACE) PCR was performed for 3' analysis of *Glyma18g02590* (Table S3 87- 90) using the SMARTer RACE cDNA kit per manufacturer protocols (ClonTech, Mountain View, CA) and previously defined primers (Cook et al., 2012). Following RACE, PCR products were TA cloned into pCR8/GW/TOPO as previously mentioned. Randomly chosen colonies were sequenced to confirm the 3' ends of individual transcripts.

4.5.3. fiber-FISH

Fiber fish experiments were carried out using the same methods and probes as previously detailed (Cook et al., 2012).

4.5.4. Whole Genome Sequencing

Whole genome sequencing was performed for lines Peking (PI 548402), PI 90763, PI 437654, PI 209332, PI 89772, and Cloud (PI 548316). Tissue was collected from at least

5 plants per sample totaling at least 3 grams of tissue to homogenize any somatic or possible intra plant DNA variants. DNA was extracted following previously published protocols (Swaminathan et al., 2007). Two separate DNA libraries were constructed for each sample. For construction of the paired-end library, DNA was randomly sheared, separated, and enriched for DNA fragments ranging from 200 bp to 400 bp in length. Adapter sequence was added to the ends of each sample for bar coding using following Illumina guidelines. Paired end libraries for samples PI 209332, PI89772, and Cloud were sampled on a single Illumina HiSeq 2000 lane producing reads of 101 bp of sequence in both directions. Paired end libraries for samples Peking, PI 90763, and PI 437654 were sequenced on Illumina's HiSeq2500 using the rapid sequencing run producing sequence of 101 bp in both the forward and reverse directions. A separate library was also constructed for each sample using larger insert sizes, known as a mate-pair library. DNA for each sample was randomly sheared, separated, and collected ranging in size from 2 kb to 3 kb. The mate-pairs libraries were constructed using the mate-pair library preparation kit from Illumina following manufactures protocols. All six libraries were sequenced in the forward and reverse direction on a single Illumina HiSeq 2000 lane generating sequencing lengths of 101 bp per direction. All samples were de-multiplexed using their respective adapter sequence and processed following Illumina's Cassava-1.8.2 pipeline to generate data in the fastq format used for downstream applications.

Sequencing for the lines in the SoyNAM project is currently forth coming (Cregan and

Diers, unpublished). Briefly, each plant sample was paired-end sequenced on an Illumina HiSeq 2000 producing reads 151 bp in length in each direction. DNA insert sizes from the samples were 450 bp.

Previously sequenced *Glycine soja* data was downloaded from the Sequenced Read Archive (SRA) section of the National Center for Biotechnology Information (NCBI), stored under accession SRA009252 (Kim et al., 2010). Data from runs SRR020188, SRR020190 and SRR020182, SRR020184 were processed for analysis in this research.

4.5.5. Short Read Genome Alignments

Rapid genome alignment for SoyNAM lines

To rapidly estimate copy number of the Rhg1 interval in the SoyNAM reads were aligned to a limited reference using the program Bowtie2 (Langmead and Salzberg, 2012). The reference for mapping was created using the Bowtie2 build indexer function with input sequence corresponding to the Williams 82 (assembly 1.89) at the *Rhg1* interval on chromosome 18 interval (1,581,000 – 1,714,000), and the homologous loci on chromosome 11 interval (37,361,000 – 37,456,000), chromosome 2 interval (47,705,000 - 47,855,000), chromosome 9 interval (45,995,000 – 46,345,000), and chromosome 14 position (4,240,265 – 4340,264). Paired-end reads were mapped using default settings. Mapped reads were processed using Samtools (Li et al., 2009), and read depth was computed using the coverageBed program of BEDtools (Quinlan and Hall, 2010) over 1kb bins ranging from 1,600,000 to 1,694,000. Read depth was estimated by summing

the number of reads corresponding to the region 5' of the Rhg1 repeat (1,600,000 – 1,631,999), the Rhg1 repeat (1,632,000 – 1,663,999), and the 3' region (1,664,000 – 1,694,000. Copy number was estimated using both flanking regions, computed as the ratio of read depth corresponding to the Rhg1 interval divided by the total reads in the flanking interval. Read depth was reported as the average of these two ratios along the standard error of the mean.

Full Genome Alignment

Illumina sequenicng reads were aligned to the full Williams 82 reference genome (build 1.89) using the program BWA (version 0.7.1) (Li and Durbin, 2009). Reads were mapped using the default settings of the *aln* function. Alignments were then paired using the *sampe* function. Alignments were further processed using the program Picard (version 1.83) to add read group information (*AddOrReplaceReadGroups)*, mark PCR duplicates (*MarkDuplicates*), and merge alignments (*MergeSamFiles*) from separate sequencing reactions per genome. For the Hg-type data processing, PCR duplicates were marked at the lane level prior to merging the sequencing runs (McKenna et al., 2010).

4.5.6. Sequence Variant Detection

Sequence alignment files were processed for variant discovery using the Genome Analysis Tool-Kit (GATK) software package (version 2.4.9) (DePristo et al., 2011). The best practices were followed as described. Insertion and deletion sites were identified using the *RealignerTargetCreator* and set list of known INDELs. Because a known

INDEL list is not publicly available for soybean, one was created following the GATK recommended guidelines. The list of known INDELs was created by selecting for concordance among high confident INDELS identified from the samples 4J105-34, LD00-3309, LG05-4292, and CL0J095-46 i.e., INDELs predicted with confidence from all 4 genomes was used as the list of knows. Following the *RealignerTargetCreator,* samples were re-aligned around INDEL sites using the *IndelRealigner* function with options:

```
 --consensusDeterminationModel USE_READS
–known INDELS
--maxConsensuses 70
--LODThresholdForCleaning 0.5
--maxReadsForConsensuses 600
--maxReadsForRealignment 100000
```

Following re-alignment, variants were called using the *UnifiedGenotyper* algorithm with options:

```
-stand_call_conf 20
-stand_emit_conf 15
-rf BadCigar
-A VariantType
-glm BOTH
```

To filter low quality variants, a filter was applied to remove variants that were not both sequenced at least three times and have a quality score greater than 50.

Variant files were annotated with the program SnpEff as documented (Cingolani et al., 2012).

4.5.7. Copy number estimates

Read depth in the 1kb intervals was averaged over the two flanking intervals to determine average read depth of the region per re-sequenced genome, and used to determine the estimated copy number of the *Rhg1* locus and the flanking intervals. We used average read depth over 1kb intervals to estimate copy number from the whole genome re-sequencing data. The analyzed interval was (93kb) centered on the known 31kb *Rhg1* repeat with equally spaced flanking intervals. The average read depth in 1kb bins was determined for the flanking *Rhg1* regions, and used to normalize read depth across bins. Final copy number estimates were made by averaging the normalized read depth across the three 32 kb intervals.

4.5.8. Phylogenetic Analysis

Phylogenetic analysis was conducted using the software package MEGA (version 5.2) (Tamura et al., 2011). The sequence for analysis included all Hg-type lines, 24 SoyNAM lines, Williams 82, *Glycine soja*, and the chromosome 11 paralog of *Rhg1* from Williams 82. The sequence included the genomic DNA of the coding sequence for Glyma18g02580, Glyma18g02590, Glyma18g02600, and Glyma18g02610 plus 1kb of genomic sequence upstream of the 5'UTR sequence. Phylogentic analysis was performed using Maximum Parsimony (MP). MP was computed using bootstrapping with 100 replicates and Close-Neighbor-Interchange on Random Trees was used with the random addition of 10 trees and a search level of 2. A consensus tree was then constructed from the possible trees using a bootstrap cutoff value of 50 to split nodes.

4.5.9. Methylation Analysis

Restriction Based Methylation Discovery

Locus specific DNA methylation was analyzed using the methylation specific endonulease McrBC, or the methylation sensitive endonuclease HpaII followed by PCR. McrBC digests DNA with methylated cytosines in a sequence independent manner while unmethylated DNA is unaffected. HpaII digest DNA at the recognition sequence CCGG, but HpaIIs endonuclease activity is blocked by cytosine methylation. Restriction digestions were performed using 600-700ng of DNA and manufacture protocols. Adding the same amount of DNA to the reaction buffer with no restriction enzyme was used to set up control reactions. Samples with and without the restriction enzyme were incubated at 37C for 90 minutes, and heat inactivated at 65C for 20 minutes. DNA was visualized in a 0.8% ethidium bromide stained gel to ensure DNA digestion. Both digested and control DNA samples were used for subsequent PCR using GoTaq flexi DNA polymerase (Promega, Madison WI). For McrBC treated DNA, PCR primers that spanned methylated DNA would not produce the intended product following PCR because the template DNA would be digested by McrBC. DNA that was not methylated or not treated with the enzyme yielded a product of the expected size. For HpaII treated DNA, PCR primers that spanned the DNA sequence CCGG in which either cytosine was methylated yielded a PCR product of the expected size. DNA sequence CCGG that was not methylated was cleaved by HpaII and failed to yield a PCR product. DNA incubated in buffer without HpaII yielded expected PCR products.

Computational Analysis

To estimate *Rhg1* copy number, the total number of cytosine sequencing reads was summed over 1kb bins starting at position 1,600,225 counting till the end of bin 1,696,224 for a total of 96 bins. Average sequencing coverage in the region was calculated by averaging the number of cytosine reads in the 1kb bins over the two 32kb intervals flanking *Rhg1*, which was used to normalize the read depth for each 1kb bin. Final copy number estimates of the three 32 kb intervals was calculated as the average normalized read depth over the respective 32kb interval. Custom scripts were written in Java or Bash to compute the data. Processed RNA-seq data was used to compare transcript levels across the 4 tested genotypes. Reads were reported in the standard reads per kilobase per million reads (RPKM) format. To assess differences, the RPKM values from the 3 replicates of the susceptible parent LDX01-1-165 were first averaged. This number was used as a normalizer for the average of the RPKM of the three replicates for the other three lines tested.

4.6. Figures

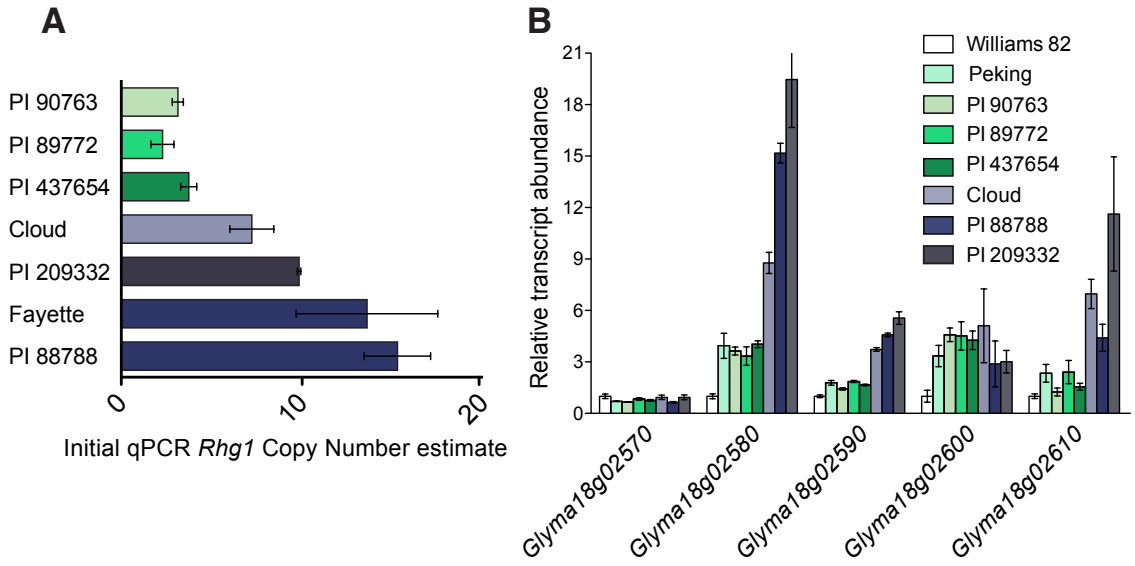


Figure 1.

Hg-type soybean lines have elevated copy number and transcript abundance for genes encoded in *Rhg1* repeat based on an initial assessment.

A. Initial copy number estimates for *Rhg1* repeats, determined using qPCR. Bars are the average copy number estimates from three replicate DNA samples along with standard error of the mean error bars. **B.** Transcript abundance estimates using qPCR for genes contained in, and adjacent (*Glyma18g02570*) to the *Rhg1* repeat. Root RNA was collected 5 days post emergence from of three independent plants grown in pots. Bars show the average transcript abundance relative to Williams 82 along with standard error of the mean error bars.

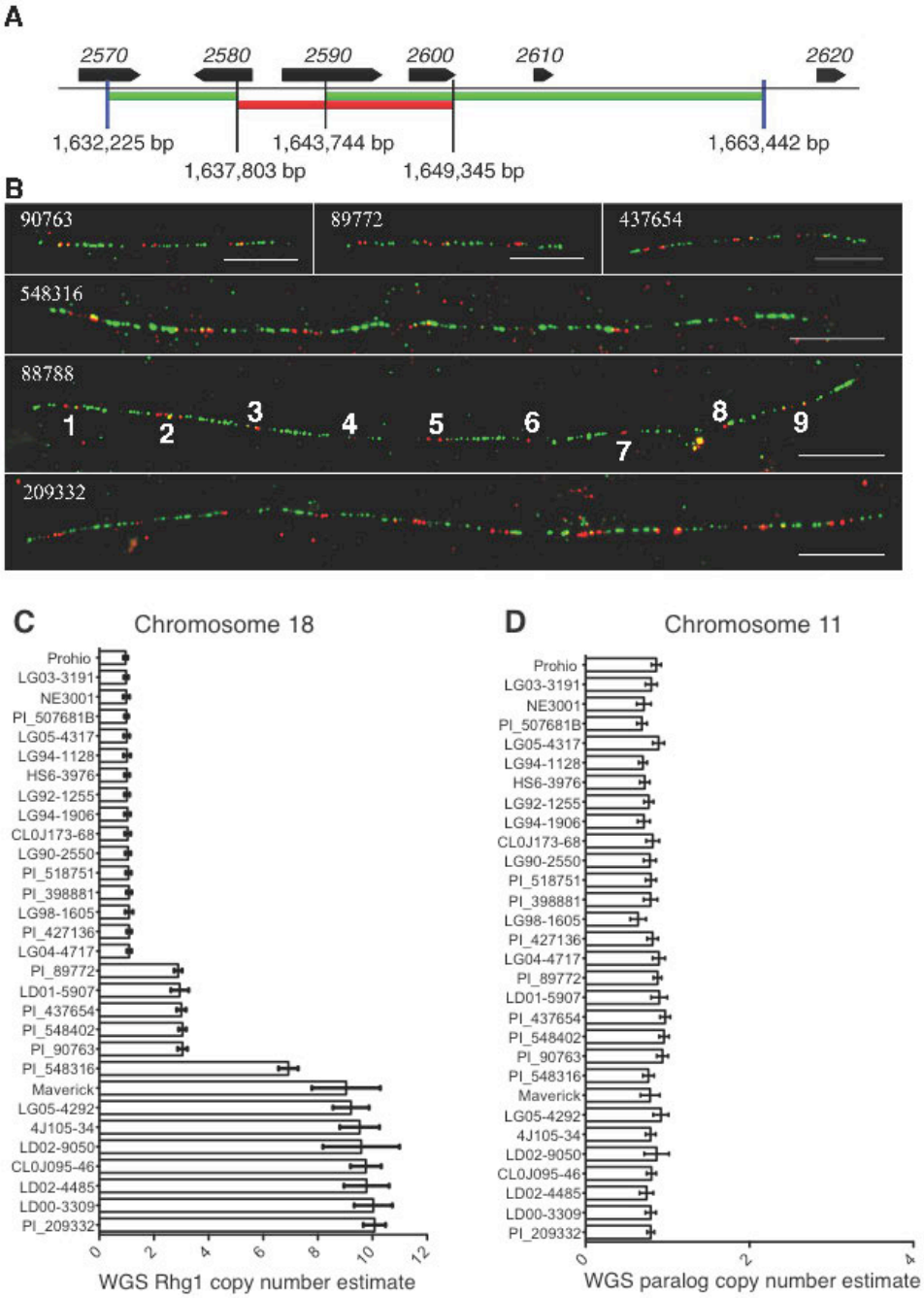


Figure 2.

Fiber-FISH and whole genome sequencing detail *Rhg1* copy number variation from a diverse set of soybean lines

A. Diagram of probe used to detect *Rhg1* repeats for fiber-FISH. **B.** Representative fiber-FISH images collected from 6 of the Hg-type soybean lines. The number of

Figure 2. cont'd

Rhg1 repeats can be determined by counting the number of times the probe (depicted in A) hybridizes to an intact DNA fiber. The name of each line is shown within the box of each image. The repeats are numbered in the box showing a representative image from PI 88788. Going from left to right, and top to bottom 3, 3, 3, 7, 9, and 10 copies of *Rhg1* shown are shown for each respective genotype. **C.** The bars show estimated *Rhg1* copy number estimates based on read depth analysis over 1kb bins across the 31kb repeat. Copy number was estimated by determining the average read depth in the two equally sized (31kb) intervals immediately adjacent to either end of the *Rhg1* repeats. This average read depth per 1kb bins was used to normalize the read depth in all 1kb bins and estimate copy number. The copy number in each 1kb bin was averaged across the three intervals (5', *Rhg1*, 3') and the estimated copy number for *Rhg1* interval is shown with standard error of the mean error bars. D. Same copy number estimates shown in panel C, but read depth is from the paralogous locus for *Rhg1* found on chromosome 11.

Figure 3.

Glyma18g02590 amino acid alignment showing polymorphic residues found in low and high-copy *Rhg1* lines

An alignment of the final 100 amino acids of α-SNAP encoded by *Glyma18g02590* shows to related but distinct polymorphic regions in low and high-copy R*hg1* classes. Positions depicted with a '.' Are identical to the consensus sequence shown from Williams 82. Residues polymorphic in at least one genotype are highlighted yellow in the consensus. Genotypes determined to be in the low-copy class are highlighted in blue, while those in the high-copy class are highlighted in green. All other genotypes contain a single *Rhg1* copy.

```
Williams     191 YGVKGHLLNAGICQLCKEDVVAITNALERYQELDPTFSGTREYRLLADIAAAIDEEDVAKFTDVVKEFDSMTPLDSWKTTLLLRVKEKLKAKELEEDDLT-
LG05-4317    191 ..............................................................................................-
LG94-1128    191 ..............................................................................................-
LG94-1906    191 ..............................................................................................-
LG92-1255    191 ..............................................................................................-
LG03-3191    191 ..............................................................................................-
LG04-4717    191 ..............................................................................................-
LG98-1605    191 ..............................................................................................-
LG90-2550    191 ..............................................................................................-
NE3001      191 ..............................................................................................-
CL0J173-86  191 ..............................................................................................-
HS6-3976    191 ..............................................................................................-
Prohio      191 ..............................................................................................-
PI 518751   191 ..............................................................................................-
PI 398881   191 ..............................................................................................-
PI 427136   191 ..............................................................................................-
PI 507681B  191 ..............................................................................................-
G.soja      191 ..............................................................................................-
Peking      191 ................E.............................................................................YEVIT
PI 90763    191 ................E.............................................................................YEVIT
PI 89772    191 ................E.............................................................................YEVIT
PI 437654   191 ................E.............................................................................YEVIT
LD01-5907   191 ................E.............................................................................YEVIT
LD00-3309   191 ............K.................................................................................QHEAIT
LG05-4292   191 ............K.................................................................................QHEAIT
LD02-4485   191 ............K.................................................................................QHEAIT
LD02-9050   191 ............K.................................................................................QHEAIT
4J105-34    191 ............K.................................................................................QHEAIT
CL0J095-46  191 ............K.................................................................................QHEAIT
Maverick    191 ............K.................................................................................QHEAIT
Cloud       191 ............K.................................................................................QHEAIT
PI 209332   191 ............K.................................................................................QHEAIT
```

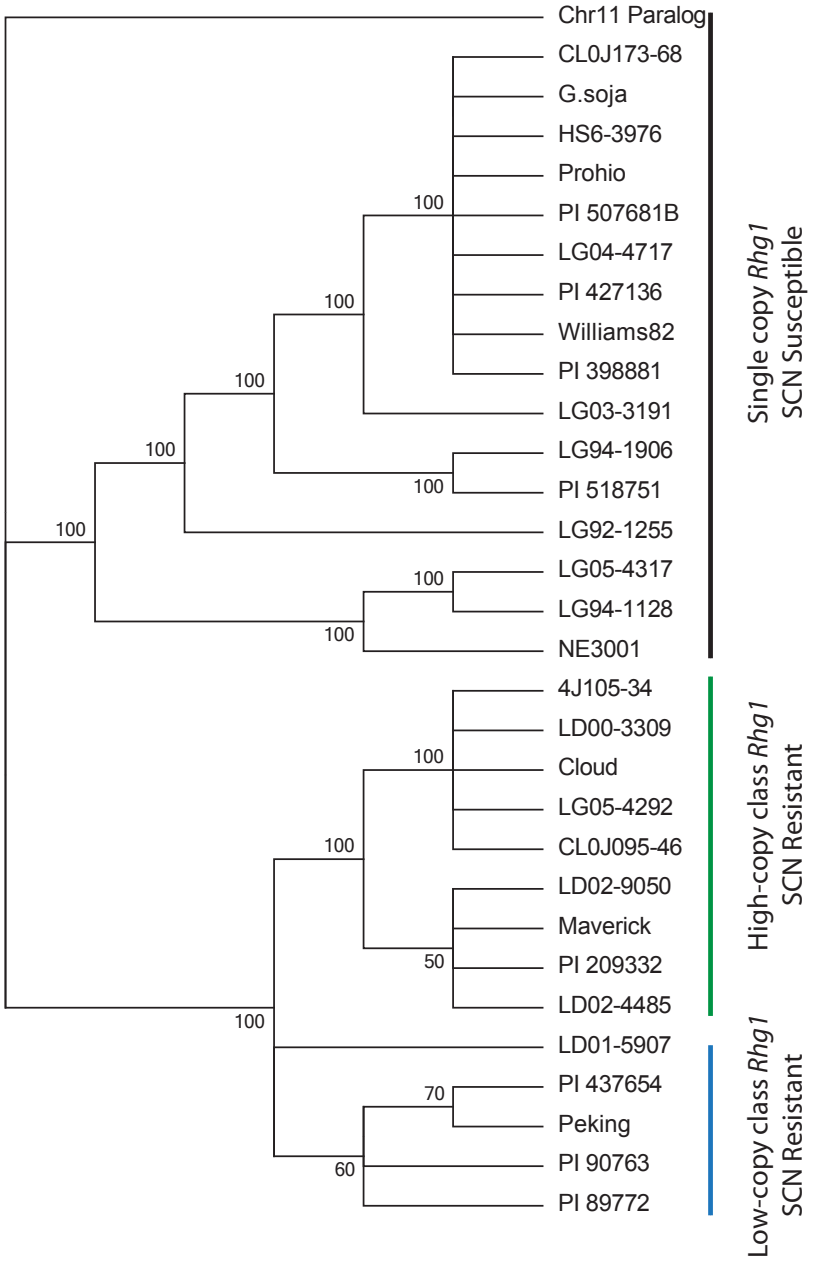


Figure 4.

Maximum parsimony phylogenetic analysis indicates a distinct split for single copy versus multi-copy *Rhg1* containing lines

Maximum parsimony phylogenetic analysis indicates a distinct split between SCN susceptible lines that contain a single copy of *Rhg1*, and SCN resistant lines

Figure 4 cont'd
containing multiple copies of *Rhg1*. An additional split is indicated between the low-copy and high-copy *Rhg1* lines. Each of the three groups are bracketed by either a black, green, or blue vertical line respectively. The phylogenetic tree was constructed by computing the consensus tree from 10 separate maximum parsimony trees using a bootstrap value of 50 to condense or split nodes. The coding sequence of *Glyma18g02580, Glyma18g02590, Glyma18g02600*, and *Glyma18g02610* plus an additional 1kb of upstream promoter for a total of 14,100bp of sequence was included for this analysis. Numbers at nodes indicate the confidence from 100 bootstrap replicates.

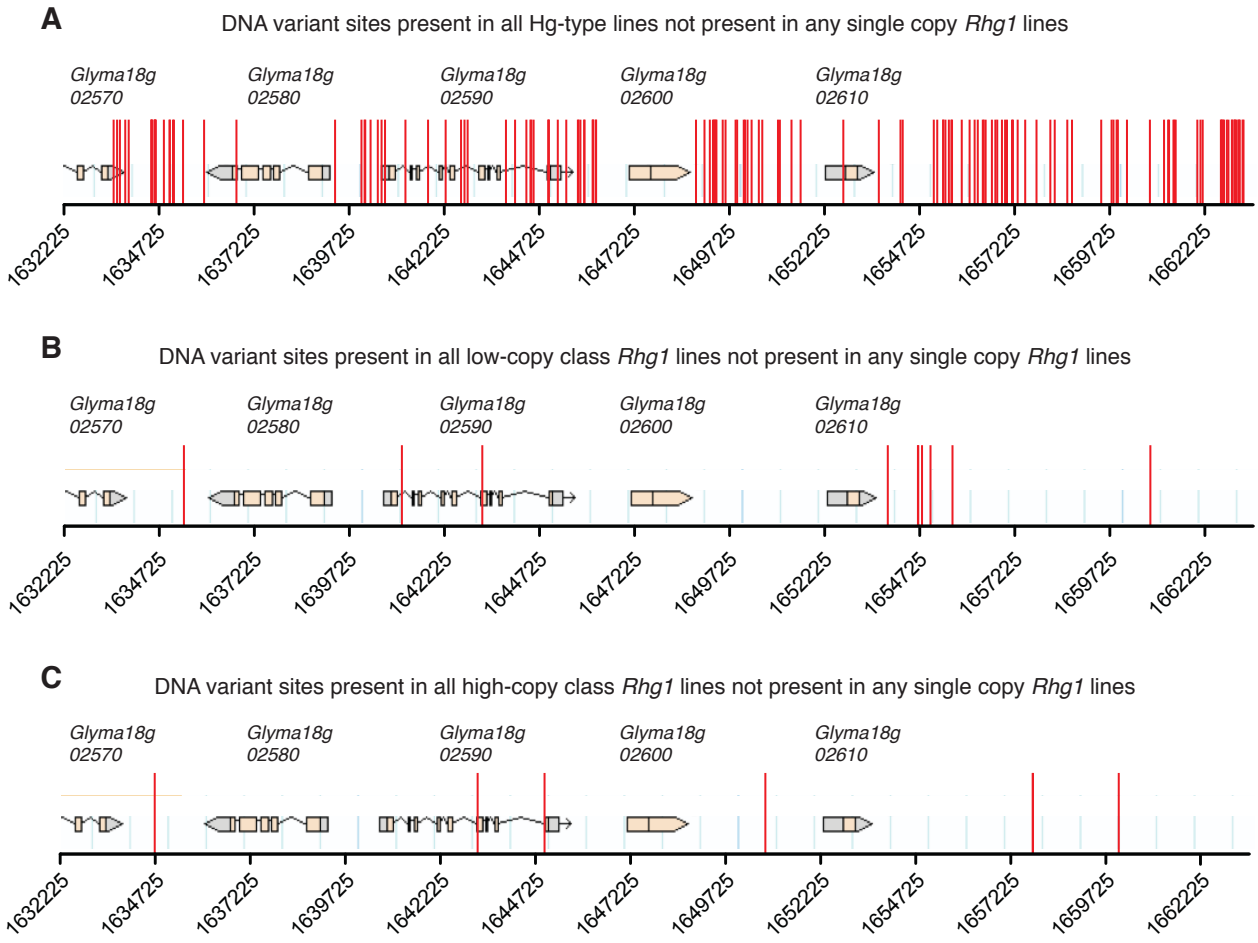


Figure 5.

Location and concordance of DNA variants sites across *Rhg1* multi-copy lines indicates a common origin

A. Schematic representation of *Rhg1* with DNA variant sites shown as vertical red lines. Gene models are shown including exons (tan boxes), UTRs (grey boxes), and introns (line segments). DNA variants were determined from whole genome sequence analysis. DNA variant sites common between SCN resistant and susceptible lines were removed, and then only the variant sites common between the 7 Hg-type lines were selected and shown **B.** Similar data as shown in A, but only DNA variants common within the low-copy class *Rhg1* lines, not present in high-copy class lines, are shown. **C.** Similar data as shown in A, but only DNA variants common within the high-copy class *Rhg1* lines, not present in low-copy class lines, are shown.

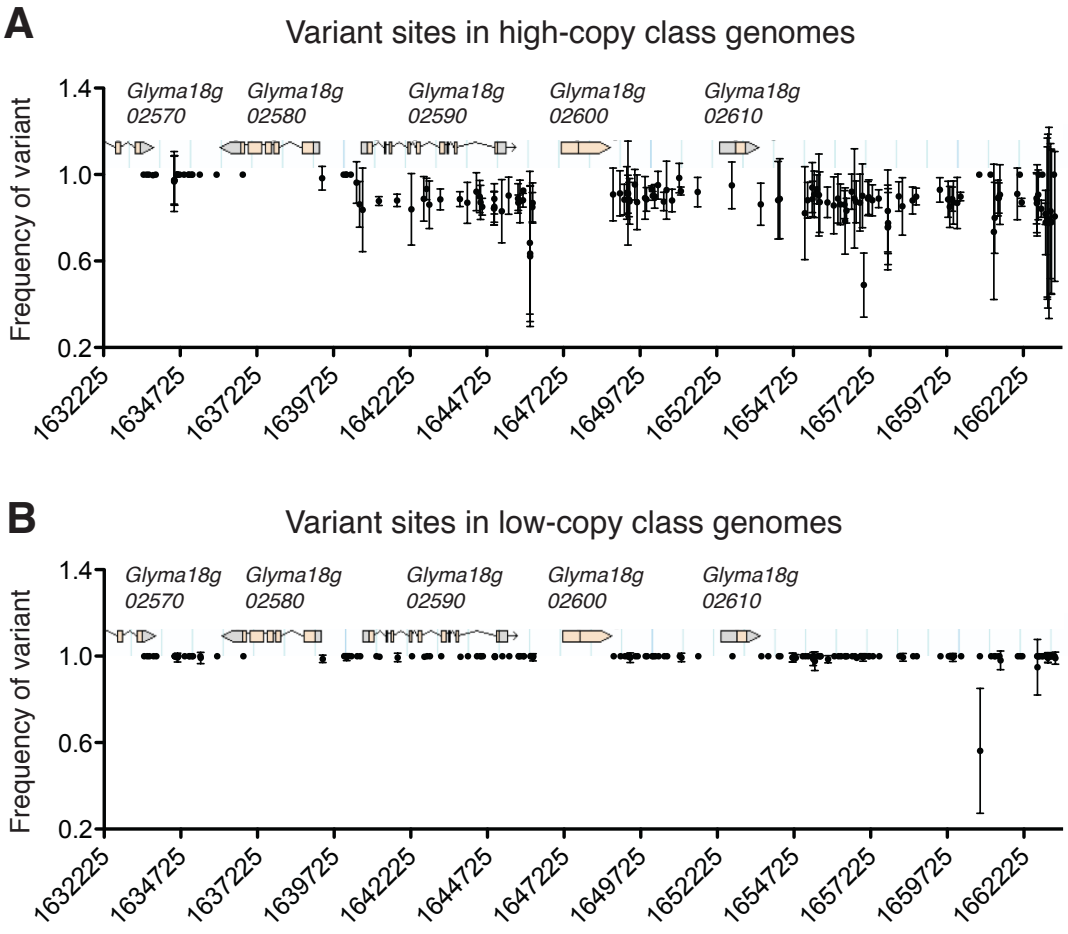


Figure 6.

Frequency of variant alleles at *Rhg1* indicates heterogeneous sequence in high-copy class but not low-copy class repeats of the *Rhg1* interval.

A. Schematic representation showing the frequency of variant sequencing reads vs reference. Each point along the x-axis indicates the site of a DNA variant, while the y-axis depicts the average frequency across the three high-copy class Hg-type lines. Only the DNA variants sites that are not found in any single-copy Rhg1 lines were included. Error bars for individual points show the standard error of the mean across the tested lines. **B.** Similar data anylsis as shown in A, but DNA variants contained in the four low-copy class Hg-type lines not present in single-copy *Rhg1* lines is shown.

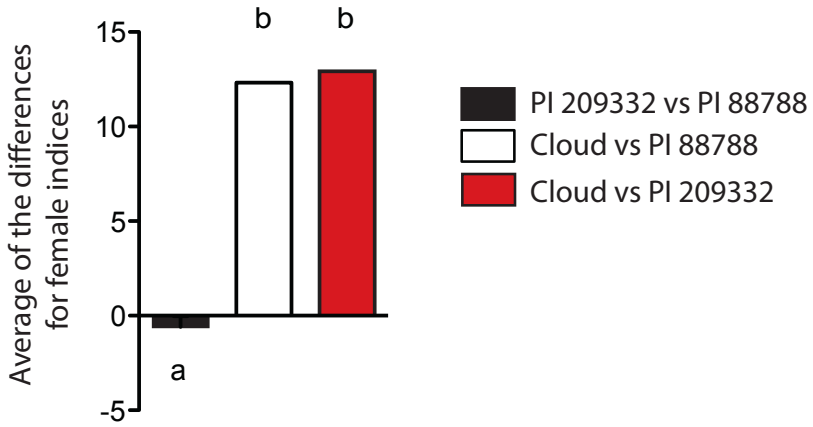
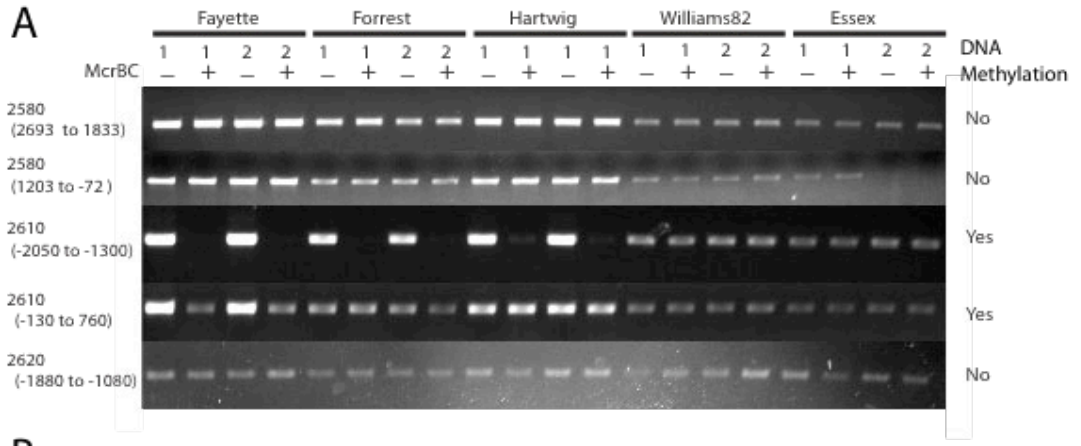


Figure 7.

Initial analysis of SCN resistance data indicates that lines containing more *Rhg1* copies from the high-copy class type may be more resistant

The average of the female index differences between indicated pairs of lines tested with the same nematode population, for data collected from 97 SCN field populations over 4 years. Data were collected as part of the northern regional SCN tests. One-way ANOVA indicates the average of the differences between Cloud and either of the other two lines is significantly different than the average of the differences between PI 88788 and PI 209332 as indicated by letters, using Tukey's post-test with p-value < 0.05.



C

Region Analyzed (nt position relative to ATG)	DNA treatment	Presence of Methylation: Fayette	Forrest	Hartwig	Williams82	Essex	Polymorphic Methylation
Gm18:2570							
-770 to -1415	McrBC	**Yes**	**Yes**		**Yes**	**Yes**	No
-115 to -850	McrBC	**Yes**	**Yes**		**Yes**	**Yes**	No
Gm18:2580							
58 to -72	McrBC & HpaII	No	No	No	No	No	No
1203 to -72	McrBC	No	No	No	No	No	No
2639 to 1833	McrBC	No	No	No	No	No	No
Gm18:2580/2590 (relative to 2590 ATG)							
-1692 to -1042	McrBC	**Yes**	No	No	No	No	Yes
-1255 to -709	McrBC	**Yes**	No		No	No	Yes
-798 to -154	McrBC	**Yes**			No		Yes
Gm18:2590							
3669 to 4230	McrBC	No	No	No	No	No	No
Gm18:2600							
-230 to 455	McrBC	No	No		No	No	No
Gm18:2610							
-2052 to -1301	McrBC	**Yes**	**Yes**	**Yes**	No	No	Yes
-1324 to -184	McrBC	**Yes**	**Yes**		No	No	Yes
-132 to 757	McrBC	**Yes**	No	No	No	No	Yes
-132 to 179	HpaII	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	No
-216 to 5	HpaII	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	No
138 to 256	HpaII	**Yes**	**Yes**	**Yes**	No	No	Yes
Gm18:2620							
-1880 to -1080	McrBC	No	No	No	No	No	No
-890 to -80	McrBC	No	No	No	No	No	No

Figure 8.
Differentially methylated root DNA at *Rhg1* occur around genes shown to impact resistance in *Rhg1* multi-copy genomes

Figure 8 cont'd

DNA of roots from SCN resistant samples Fayette (PI 88788 derived resistance), Forrest (Peking derived resistance), and Hartwig (PI 437654 derived resistance), along with the SCN susceptible lines Williams 82 and Essex were analyzed for DNA methylation. A,B. Representative PCR products using DNA either treated (+) or mock (-) treated with the methylation sensitive restriction enzyme McrBC or HpaII. Following DNA treatment, PCR was conducted using the primers spanning the indicated region in the left column. Two independent DNA samples of roots were collected and analyzed from separate plants, indicated by a 1 or 2. **A.** Intensity of PCR products amplified from DNA treated (+) with McrBC are expected to decrease compared to those amplified from mock (-) treated DNA if the primers span methylated cytosines. **B.** PCR products spanning CCGG sites only amplify following HpaII digestion if digestion is blocked by methylation of the inner cytosine. **C.** Summary results from 18 primer pairs spanning various regions of the *Rhg1* locus. The presence of methylation is listed along with regions where differential methylation was detected between a SCN resistant and susceptible genome. Blank cells were not analyzed.

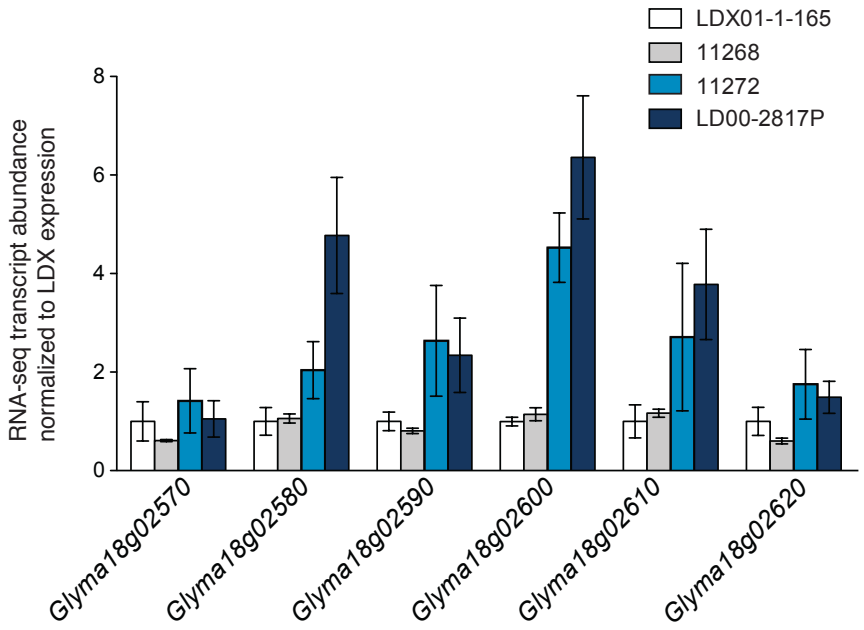


Figure 9.

Lines estimated to contain 3 copies of the *Rhg1* interval show increased expression for the genes encoded in the repeat based on RNA-seq analysis

The parental line LD00-2817P and progeny line 11272 are estimated to contain roughly three *Rhg1* copies (Table 4). These two lines display higher expression of the genes encoded in the interval relative to the parental line LDX01-1-165 and progeny line 11268 estimated to contain a single copy of *Rhg1*. RNA-sequencing is reported in reads per kilobase per million reads and normalized to expression from line LDX01-1-165.

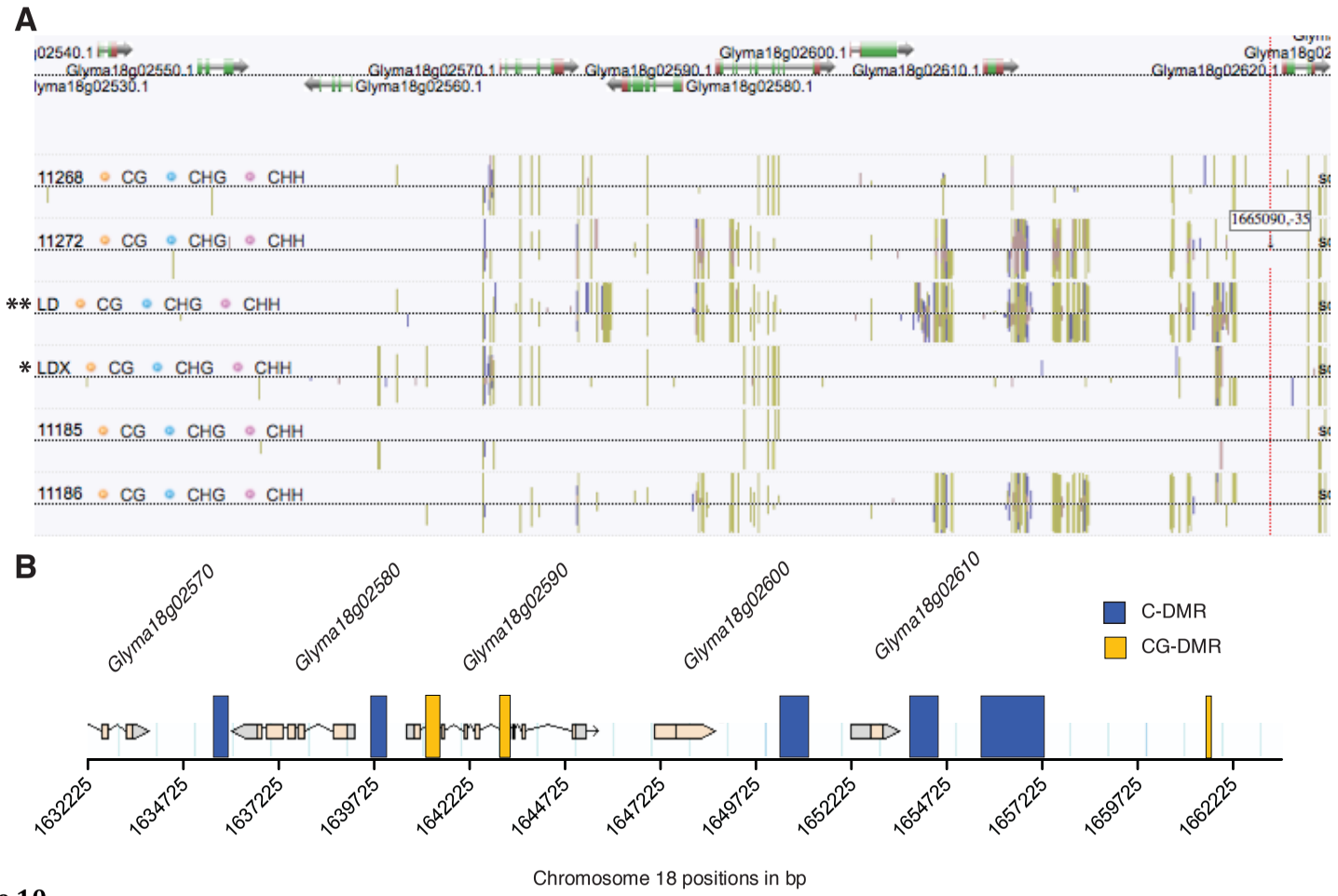


Figure 10.

Significant differential methylation surrounding genes shown to impact *Rhg1*-mediated SCN resistance

A. Genome browser depicting DNA methylation level and position as vertical lines. The height of the line is proportional to the level of methylation. Bar colors represent the sequence context of the cytosine methylation as indicated by the key at the left side of each line. Each line corresponds to a different genotype analyzed with the parental lines containing either a single copy (*) or three copies (**) of *Rhg1*denoted respectively. **B.** Summary of C-DMR and CG-DMR identified Schmitz et al., 2013. Colors indicate the sequence context of the DMR.

Table 1.

DNA variant summary, including SNPs, insertions and deletions from diverse soybean germplasm

Summary statistics for the different types of DNA variants discovered, and summary totals for the genomic location of the variants shown for each genotype. * denotes copy number determined via fiber-FISH.

Genotype	*Rhg1* Copy Number	SNP			Insertion			Deletion			Genome Location			
		Total	Homo	Hetero	Total	Homo	Hetero	Total	Homo	Hetero	Exon	Intron	Intergenic	5', 3' UTR
LD00-3309	10.03	192	129	63	36	14	22	34	15	19	6	24	224	8
LG05-4292	9.23	197	124	73	38	19	19	34	18	16	6	25	230	8
4J105-34	9.53	196	143	53	35	14	21	34	17	17	6	24	227	8
CL0J095-46	9.76	195	139	56	36	18	18	34	17	17	6	24	227	8
LD02-4485	9.79	191	139	52	36	22	14	33	17	16	6	25	221	8
LD02-9050	9.56	177	113	64	32	17	15	28	18	10	6	24	199	8
Maverick	9.04	187	126	61	33	16	17	31	16	15	6	24	213	8
Cloud	7*	194	94	100	35	14	21	34	14	20	6	25	224	8
PI 209332	10*	197	159	38	37	17	20	33	17	16	6	27	226	8
PI 437654	3*	193	190	3	38	35	3	36	33	3	5	27	226	9
PI 548402	3*	200	195	5	40	38	2	35	32	3	6	25	236	8
PI 89772	3*	196	189	7	37	37	0	36	34	2	5	25	231	8
PI 90763	3*	192	189	3	38	37	1	35	33	2	5	25	226	9
LD01-5907	2.95	146	135	11	18	17	1	17	17	0	5	24	146	6
NE3001	0.99	63	63	0	3	3	0	6	6	0	4	23	42	3
LG05-4317	1.01	72	69	3	4	4	0	5	5	0	1	19	60	1
LG94-1128	1.02	45	45	0	2	2	0	3	3	0	1	15	33	1
LG92-1255	1.02	46	42	4	6	5	1	3	3	0	0	3	51	1
PI 518751	1.07	44	44	0	2	1	1	3	3	0	0	0	49	0
LG94-1906	1.03	26	1	25	0	0	0	2	1	1	0	0	28	0
LG03-3191	0.98	28	1	27	0	0	0	0	0	0	0	0	0	0
CL0J173-68	1.04	0	0	0	0	0	0	0	0	0	0	0	0	0
HS6-3976	1.02	0	0	0	0	0	0	0	0	0	0	0	0	0
LG04-4717	1.10	0	0	0	0	0	0	0	0	0	0	0	0	0
PI 398881	1.09	0	0	0	0	0	0	0	0	0	0	0	0	0
PI 427136	1.09	0	0	0	0	0	0	0	0	0	0	0	0	0
PI 507681B	1.00	0	0	0	0	0	0	0	0	0	0	0	0	0
LG90-2550	1.05	0	0	0	0	0	0	0	0	0	0	0	0	0
Prohio	0.97	0	0	0	0	0	0	0	0	0	0	0	0	0
LG98-1605	1.09	0	0	0	0	0	0	0	0	0	0	0	0	0
G.soja	1.10	42	42	0	3	3	0	4	4	0	1	3	45	0

Table 2.

Amino acid substitutions for genes encoded within and adjacent to the *Rhg1* interval showing two new alleles of α-SNAP in multi-copy *Rhg1* lines

Amino acid substitutions are grouped by genes and listed with the variant base pair position resulting in the substitution. Variants are reported as the amino acid found in Williams 82, the amino acid position in the full-length protein, and the new amino acid at that position. Amino acid substitutions were the same in the low and high-copy class lines and reported together. The one exception was a nucleotide deletion in the open reading frame of *Glyma18g02600* indentified only in Peking. This deletion offset the predicted coding frame resulting in a stop codon 8 amino acids later. Amino acid substitutions were only detected in three other genotypes and are reported in the table. All other single-copy *Rhg1* lines have no detected amino acid substitutions. No amino acid substitutions were detected in any genome for the genes highlighted in grey.

Position (bp)	Low-Copy Lines	High-Copy Lines	NE3001	LG05-4317	LG94-1128
Glyma18g02570 (unknown function)					
Glyma18g02580 (amino acid transporter)					
1638989			V9I		
Glyma18g02590 (alpha-SNAP)					
1643208		Q203K			
1643225	D208E				
1644965		E285Q			
1644968	D286Y	D286H			
1644972	D287EV	D287EA			
1644974	L288I	L288I			
Glyma18g02600 (PLAC-8 family)					
1647134			V23A		
1647764	-214 (Peking)				
1648561				M480L	M480L
Glyma18g02610 (unknown function)					
Glyma18g02620 (SEL-1-like)					

Table 3.

Correlations for female index values indicates significant groupings among multi-copy *Rhg1* containing Hg-type lines

Pearson correlation coefficients between soybean cyst nematode resistant lines computed from female indices as part of the northern regional soybean cyst nematode tests conducted from 2009 to 2012. A high correlation coefficient indicates, for example, that across many nematode populations, the extent of resistance of Peking plant to any given nematode population tended to be similar to that of PI 90762 for the same nematode population. r^2 values are reported along with significance corresponding to two different p-value cutoffs indicated by colors.

Genotype	Peking	PI 88788	PI 90763	PI 437654	PI 209332	PI 89772
PI 88788	0.45					
PI 90763	0.65	0.04				
PI 437654	0.08	0.03	0.11			
PI 209332	0.46	0.88	0.05	0.02		
PI 89772	0.79	0.36	0.66	0.02	0.34	
Cloud	0.34	0.76	0.04	0.06	0.81	0.27
Statistical significance for r^2 displayed as			$p < 0.05$	$p < 0.001$		

Table 4.

Copy number estimates for *Rhg1* and the equally sized adjacent intervals indicates that parental line LD00-2817P and some progeny contain multiple copies of *Rhg1*

The genotype column lists the names of the genotypes with parental line LDX01-1-165 and LD00-2817P abbreviated as LDX and LD respectively. The coverage column indicates the average estimated sequence coverage based on whole genome sequencing. Copy number was estimated for the 31kb *Rhg1* repeat interval, along with the equally sized intervals positioned at the 5' and 3' ends. Read depth across the intervals was analyzed in 1kb bins. Read depth across the two flanking intervals was averaged and used to normalize read depth and estimate copy number for each 1kb bin. Estimated copy number was determined by averaging this copy number estimate over the 3 independent intervals.

		Estimated Copy Number		
Coverage	**Genotype**	**5' to repeat**	***Rhg1***	**3' to repeat**
5.2	11246	1.12	1.00	0.88
4.0	11211	1.12	1.02	0.88
4.2	11240	1.15	1.04	0.85
4.3	11210	1.14	1.07	0.86
4.0	11230	1.11	1.07	0.89
4.6	11251	1.10	1.09	0.90
4.1	11225	1.14	1.10	0.86
4.3	11232	1.11	1.10	0.89
7.0	11221	1.16	1.11	0.84
4.2	11267	1.11	1.11	0.89
4.0	11259	1.08	1.12	0.92
4.4	11248	1.24	1.12	0.76
4.0	11247	1.09	1.13	0.91
12.9	11209	1.16	1.15	0.84
13.8	11237	1.10	1.16	0.91
17.7	LDX	1.07	1.17	0.94
10.9	11268	1.16	1.24	0.84
5.4	11253	1.06	1.31	0.94
11.9	11217	1.24	1.38	0.76
7.5	11234	1.05	1.45	0.95
5.2	11239	1.10	1.51	0.90
10.6	11218	1.10	1.99	0.90
4.0	11228	1.02	2.52	0.99
14.3	11249	1.02	2.54	0.98
4.0	11262	1.00	2.55	1.00
4.1	11186	1.00	2.68	1.00
4.3	11276	1.02	2.71	0.98
11.6	11272	1.07	2.85	0.93
16.9	LD	0.98	2.93	1.02

4.8. Supplemental Figures

```
Ch18 Williams          421 TTTGCTGCTCAGCTAGAACAATATCAGAAGTCGATTGACATTTATGAAGAGATAGCTCGC
Chr18 PI 88788         421 ............................................................
Chr18 Peking           421 ............................................................
Peking_isoform         421 ............................................................
Chr11 Williams         421 ..........................................G.....C............
Truncated α-SNAP       421 ..........................................G.................
Chr11 Peking           421 ..........................................G.....C.........A.....


Ch18 Williams          481 CAATCCCTCAACAATAATTTGCTGAAGTATGGAGTTAAAGGACACCTTCTTAATGCTGGC
Chr18 PI 88788         481 ............................................................
Chr18 Peking           481 ............................................................
Peking isoform         481 ............................................................
Chr11 Williams         481 ................................................G...........
Truncated α-SNAP       481 ............................................................
Chr11 Peking           481 ................................................G...........

Ch18 Williams          541 ATCTGCCAACTCTGTAAAGAGGACGTTGTTGCTATAACCAATGCATTAGAACGATATCAG
Chr18 PI 88788         541 ......A.....................................................
Chr18 Peking           541 ......................G.....................................
Peking isoform         541 ......................G.-----------------------------------
Chr11 Williams         541 ...................G...T...A.....G.........................
Truncated α-SNAP       541 ...................G.......A.....G.........................
Chr11 Peking           541 ...................G...T...A.....G.........................

Ch18 Williams          601 GAACTGGATCCAACATTTTCAGGAACACGTGAATATAGATTGTTGGCGGACATTGCTGCT
Chr18 PI 88788         601 ............................................................
Chr18 Peking           601 ............................................................
Peking isoform         566 -...........................................................
Chr11 Williams         601 ......................G.....................T...............
Truncated α-SNAP       601 ..........................................T......TTAGG.CACTAG
Chr11 Peking           601 ..........................................T......TTAGG.CACTAG

Ch18 Williams          661 GC
Chr18 PI 88788         661 ..
Chr18 Peking           661 ..
Peking isoform         625 ..
Chr11 Williams         661 ..
Truncated α-SNAP           --
Chr11 Peking               --
```

Supplemental Figure 1.

Alternative transcripts of *Glyma18g02590* encoded at *Rhg1* in SCN resistant lines, and an alternative transcript for the *Glyma18g02590* paralog

Nucleic acid alignment for the final 240 bases of transcript sequence from Williams, PI 88788 and Peking, along with the sequence of the paralog encoded on chromosome 11 from Williams, Peking, and the sequence reported in (Matsye et al., 2012). Positions with a . have the same sequence as that shown for Williams 82, and the truncated sequence identified in Peking is marked by "-". The sequence reported for the truncated allele of *Glyma18g02590* is most similar to the Williams 82 and Peking paralog encoded in chromosome 11. Peking sequenced identified based on whole genome sequence analysis. Polymorphisms affecting transcript splicing and resulting in a new stop codon are highlighted in yellow and red respectively.

4.9. Supplemental Tables

Supplemental Table 1.

Summary statistics for whole genome sequencing

Summary data for whole genome sequencing from **A.** SoyNAM **B.** Hg-type lines **C.** *Glycine soja.* The total amount of sequence generated is detailed in terms of the number of sequences generated, the length of the sequencing reads, and the resulting total length of sequence generated, reported in megabases. The average genome coverage is reported based on the size of the assembled soybean chromosomes and the amount of sequence generated from each library for each genotype. Mean phred quality scores for the sequenced bases is reported for data in A and B. All sequencing from A was performed on paired-end libraries of each genotype.

A

Genotype	Sequence in Megabases	Average Genome Coverage	Number of Reads	Mean Quality Score	Read Length (bp)
LD00-3309	12,443	13.1	82,405,208	34.3	151
LG05-4292	11,725	12.3	77,646,454	31.05	151
4J105-3-4	12,112	12.7	80,213,570	32.9	151
CL0J095-4-6	9,641	10.1	63,847,612	32.58	151
LD02-4485	7,565	8.0	50,099,482	34.48	151
LD02-9050	6,191	6.5	40,999,452	34.19	151
Maverick	5,750	6.1	38,077,200	34.67	151
LD01-5907	5,915	6.2	39,174,146	31.51	151
LG05-4317	9,464	10.0	62,678,462	30.94	151
NE3001	8,535	9.0	56,524,252	34.51	151
PI518_751	14,566	15.3	96,464,146	31.88	151
LG92-1255	9,227	9.7	61,105,816	31.78	151
LG94-1128	6,025	6.3	39,903,618	31.93	151
LG94-1906	8,126	8.6	53,811,902	32.95	151
CL0J173-6-8	6,375	6.7	42,218,296	32.53	151
HS6-3976	9,026	9.5	59,774,700	32.4	151
LG03-3191	11,802	12.4	78,160,256	31.49	151
LG04-4717	9,997	10.5	66,208,394	31.23	151
PI398_881	10,717	11.3	70,971,440	29.14	151
PI427_136	13,702	14.4	90,738,654	28.98	151
PI507_681B	11,330	11.9	75,029,874	29.56	151
LG90-2550	5,441	5.7	36,035,936	31.5	151
Prohio	10,322	10.9	68,354,318	32.39	151
LG98-1605	5,344	5.6	35,391,702	32.82	151

Supplemental Table 1 cont'd

B

Genotype	Sequence in Megabases	Average Genome Coverage	Number of Reads	Mean Quality Score	Read Length (bp)
Cloud					
Paired-end	11,860	12.5	117,426,128	34.98	101
Mate-pair	1,944	2.0	19,252,220	34.12	101
PI 209332					
Paired-end	14,314	15.1	141,724,130	35.16	101
Mate-pair	2,994	3.2	29,642,412	33.93	101
PI 437654					
Paired-end	17,519	18.4	173,457,140	35.48	101
Mate-pair	6,205	6.5	61,433,454	34.07	101
Peking					
Paired-end	45,670	48.1	452,180,560	35.65	101
Mate-pair	1,756	1.8	17,390,516	34.71	101
PI 89772					
Paired-end	12,894	13.6	127,662,706	35.23	101
Mate-pair	4,247	4.5	42,049,176	33.86	101
PI 90763					
Paired-end	17,043	17.9	168,742,468	35.46	101
Mate-pair	3,002	3.2	29,720,314	34.15	101

C

Genotype	Sequence in Megabases	Average Genome Coverage	Number of Reads	Read Length (bp)
Glycine soja				
Paired-end	7,573	8.0	210,379,574	36
Paired-end	17,891	18.8	232,353,078	77

Genotype	Copy Number Estimate	
	Chromosome 18 (*Rhg1*)	Chromosome 11 (Paralog)
4J105-3-4	9.9 ± 1.9	1.0 ± 0.2
LD00-3309	9.9 ± 1.8	0.9 ± 0.2
LD02-4485	9.8 ± 2.2	1.0 ± 0.3
CL0J095-46	9.6 ± 1.5	0.9 ± 0.2
LD02-9050	9.4 ± 3.4	1.0 ± 0.4
LG05-4292	9.4 ± 1.7	1.0 ± 0.2
Maverick	9.2 ± 3.3	0.9 ± 0.3
LD01-5907	2.9 ± 0.9	1.1 ± 0.3
PI574_486	1.3 ± 0.2	
LG05-4317	1.3 ± 0.2	
LG97-7012	1.2 ± 0.1	
LG04-6000	1.2 ± 0.1	
LG04-4717	1.2 ± 0.1	
LG98-1605	1.1 ± 0.4	
PI_427136	1.1 ± 0.3	
PI_404188A	1.1 ± 0.3	
LG90-2550	1.1 ± 0.3	
U03-100612	1.1 ± 0.2	
PI_398881	1.1 ± 0.2	
5M20-252	1.1 ± 0.2	
S06-13640	1.1 ± 0.2	
LG05-4832	1.1 ± 0.1	
LG94-1906	1.1 ± 0.1	
CL0J173-68	1.1 ± 0.1	
LG94-1128	1.1 ± 0.1	
PI_518751	1.0 ± 0.3	
LG92-1255	1.0 ± 0.3	
HS6-3976	1.0 ± 0.3	
Prohio	1.0 ± 0.2	
PI_561370	1.0 ± 0.2	
PI_507681B	1.0 ± 0.2	
LG03-3191	1.0 ± 0.2	
LG03-2979	1.0 ± 0.2	
IA3023	1.0 ± 0.2	
NE3001	0.9 ± 0.3	
LG05-4464	0.9 ± 0.2	

Supplemental Table 2.

Estimated *Rhg1* copy number for SoyNAM lines using rapid mapping

Short reads from whole genome sequencing were aligned to a portion of the reference genome to rapidly estimate *Rhg1* copy number. The resulting copy number estimate is reported along with the standard error of the mean. Copy number was estimated by summing the total number of reads in three equally sized DNA intervals spanning the *Rhg1* repeat, and the 5' and 3' adjacent intervals. The total number of reads from the *Rhg1* interval was independently divided by the total number of reads from the two adjacent intervals to estimate copy number. The reported copy number is the average of the two estimates along with the standard error of the mean. Copy numbers highlighted in blue are estimated to be in the high-copy class, green in the low-copy class, and all other lines were estimated to contain a single copy of *Rhg1*. Similar analysis was performed for the paralogous sequence on chromosome 11, and did not significantly deviate from a single copy.

Sequenced allele	cDNA	
	# Sequenced	Percentage
PI 88788		
High copy allele	26	0.93
Low copy allele	0	0.00
Truncated isoform	0	0.00
Williams-type	2	0.07
PI 209332		
High copy allele	8	1.00
Low copy allele	0	0.00
Truncated isoform	0	0.00
Williams-type	0	0.00
Cloud		
High copy allele	7	0.88
Low copy allele	0	0.00
Truncated isoform	0	0.00
Williams-type	1	0.12
Peking		
High copy allele	0	0.00
Low copy allele	8	0.89
Truncated isoform	1	0.11
Williams-type	0	0.00
PI 90763		
High copy allele	0	0.00
Low copy allele	6	0.67
Truncated isoform	3	0.33
Williams-type	0	0.00
PI 89772		
High copy allele	0	0.00
Low copy allele	6	0.86
Truncated isoform	1	0.14
Williams-type	0	0.00
PI 437654		
High copy allele	0	0.00
Low copy allele	8	0.73
Truncated isoform	3	0.27
Williams-type	0	0.00
Williams 82		
High copy allele	0	0.00
Low copy allele	0	0.00
Truncated isoform	0	0.00
Williams-type	6	1.00

Supplemental Table 3.

Sequenced cDNA products confirms the expression of multiple alleles of *Glyma18g02590* in the different multi-copy *Rhg1* classes

The number and percentage of sequenced cDNA products corresponding to different alleles of *Glyma18g02590* from the Hg-type lines and Williams 82. A zero indicates the transcript was not observed. High-copy lines PI 88788 and Cloud encode a Williams 82-type *Glyma18g02590* at a low frequency, not detected in low-copy genomes. Although the Williams 82-type allele was not detected in the limited cDNA sequencing of genotype PI 209332, whole genome analysis of the DNA indicates the sequence is present. A splice isoform was detected in all low-copy genomes not detected in Williams 82 or the high-copy genomes. In total, both multi-copy classes express two different alleles of

Supplemental Table 4.

Average sequence variant frequencies at positions across the *Rhg1* repeat indicate varying sequence content between copies.

Position	AlleleFreq
1633532	1
1633629	1
1633700	1
1633840	1
1633930	1
1634533	1
1634534	1
1634535	1
1634536	1
1634610	1
1634620	1
1634626	1
1634635	1
1634643	1
1634856	1
1635001	1
1635014	1
1635093	1
1635120	1
1635364	1
1635912	1
1636766	1
1639354	1
1640056	1
1640137	1
1640151	1
1640292	1
1640480	0.929
1640581	0.929
1640675	0.5
1640675	0.286
1641208	1
1641800	0.857
1642266	0.857
1642672	0.929
1642762	0.786
1642848	0.857
1643849	0.929
1644089	0.857
1644385	0.857
1644493	0.929
1644525	0.857
1644577	0.786
1644968	0.214
1644968	0.571
1644972	0.214
1644972	0.571
1644974	0.857
1645218	0.857
1645437	0.786
1645745	0.929
1645759	0.786
1645811	0.857
1645908	0.857

Position Cont'd	AlleleFreq Cont'd
1645914	0.786
1646130	0.786
1646138	0.714
1646138	0.071
1646145	0.786
1646211	0.786
1646226	0.786
1648850	0.929
1649069	1
1649212	0.929
1649293	0.929
1649328	0.929
1649335	0.786
1649371	0.929
1649385	0.929
1649553	0.786
1649630	0.857
1649892	0.929
1649934	0.857
1650106	0.786
1650140	1
1650201	1
1650310	0.786
1650501	0.929
1650590	0.857
1651003	1
1651056	1
1651364	0.357
1651364	0.429
1651602	0.786
1652723	1
1653661	0.857
1654230	0.786
1654282	0.786
1655099	0.786
1655195	0.857
1655348	0.857
1655353	0.857
1655408	0.929
1655500	0.857
1655564	0.857
1655585	0.929
1655836	0.929
1656044	0.929
1656178	0.786
1656263	0.857
1656394	0.857
1656417	0.786
1656462	0.786
1656633	0.857
1656719	0.857
1656769	0.929
1656898	0.929
1656979	0.929

Position Cont'd	AlleleFreq Cont'd
1657025	0.786
1657162	0.929
1657183	0.929
1657307	0.857
1657506	1
1657803	0.714
1658170	0.929
1658284	0.857
1658617	0.929
1658735	0.786
1659502	0.857
1659777	0.857
1659829	0.857
1659914	0.929
1659945	0.786
1660183	1
1660790	1
1661155	1
1661264	0.786
1661293	0.786
1661406	0.786
1661428	0.929
1661460	0.929
1662031	0.857
1662115	1
1662177	0.786
1662656	0.786
1662666	0.929
1662682	0.929
1662714	0.786
1662734	1
1662810	0.786
1662851	1
1662946	0.786
1662953	0.786
1663007	0.857
1663014	0.857
1663032	0.857
1663064	0.857
1663114	0.857
1663133	0.857
1663148	0.857
1663225	1
1663250	0.857
Average	0.86

Supplemental Table 4 cont'd

The average variant allele frequency at each base pair position along chromosome 18 is reported for PI 209332, Cloud, and LD00-3309– the high-copy class genomes. Variant allele frequency is calculated by summing the total number of reads supporting an alternate sequence at the position, divided by the total number of sequenced reads (wild-type plus variant) at the position. The average over all sites is shown at the end of the final frequency column, and indicates that most variant alleles are present in most of the copies.

5. Conclusions and Future Directions

5.1. Conclusions

The work reported in this thesis marks a substantial increase in our knowledge of *Rhg1*-mediated SCN resistance, building on decades of previous research. We now know that multiple, tightly clustered genes controls *Rhg1*, each required for full resistance as determined by gene silencing and complementation experiments in our transgenic root system. We have also determined that the DNA encoding these genes is present in multiple tandem copies resulting in increased expression. We have identified that the two common types of SCN resistance, referred to as PI 88788 or Peking-type, make up two different classes of *Rhg1*, distinguishable in terms of copy number, *Glyma18g02590* coding sequence, and transcript abundance. The PI 88788-type resistance, referred to here as the high-copy class including Cloud and PI 209332, appears to have a more variable copy number than the low-copy class, Peking-type resistance, including PI 90763, PI 89772, and PI 437654. The presence of distinct alternate alleles of the gene encoding a putative α-SNAP in the two multi-copy classes is also a noteworthy finding given the relatively invariant nature of this gene from a variety of plant species. Preliminary analysis of SCN resistance phenotypes, previously collected, indicates that having additional copies of *Rhg1* may provide more resistance in the high-copy class type resistance. The absence of variation in *Rhg1* copy number in the low-copy class type resistance precludes this analysis. We have also identified the presence of hypermethylated

DNA segments within the *Rhg1* not present in analyzed SCN susceptible genomes. Strikingly, the differentially methylated regions occur within and adjacent to the three genes shown to impact *Rhg1*-mediated resistance. These results are further confirmed by the experiments of Schmitz et al., 2012 and further re-analysis expands these findings. Despite the advances presented in this thesis, a number of important and interesting questions remain to be investigated, explored below in the form of 21 future questions.

5.2. Future Directions

The findings reported in this thesis have raised numerous new research questions. There are perhaps too many questions to fully describe, but the following short-form list may be of interest to those who study the topic in the future:

1. Why are three gene needed for full resistance?
2. Do the three genes act in the same pathway, different but converging pathways, or independently?
3. What is the role of the alternate α-SNAP sequences for nematode resistance?
4. How would stacking the low and high-copy class alleles of α-SNAP impact SCN resistance?
5. Does the truncated isoform of α-SNAP in low-copy lines have a functional significance?
6. Do more copies of *Rhg1* result in increased SCN resistance?
7. Would multiple copies of *Rhg1* of the low-copy class type function in the absence of *Rhg4*?
8. Would an equal number of the two types of *Rhg1* (i.e. 8 copies of the low and high copy-type) in the same genetic background offer similar levels of resistance to a similar range of SCN populations?
9. Is there variation in copy number within the low-copy class lines, or is three the maximum?
10. If there are no natural occurrences of greater than three *Rhg1* repeats of the low-copy class type, why?
11. Are there additional coding alleles for any of the genes within the *Rhg1* repeat in additional *Glycine max* accessions, or other *Glycine* species?
12. Does *Rhg1* function in alternate host for resistance to additional nematodes or pathogens?
13. Does DNA methylation at *Rhg1* directly impact SCN resistance?
14. Is the DNA methylation at *Rhg1* dynamic in response to SCN infection, and if so, what is the signal that triggers the change? (see question 15)
15. If the genes are up-regulated following nematode attack, what is the signal, receptor, and/or pathway leading to their detection?
16. What is the significance of the copia-like LTR at the 3' boarder of the repeats?
17. What is the evolutionary relationship between the low and high-copy classes of resistance?
18. How common is the control of quantitative traits by copy number variable loci?
19. How common is the control of adaptive traits by copy number variable loci?
20. How common is differential methylation at copy number variable loci?
21. Could "induced" segmental duplication be used in conjunction with plant breeding to develop new genetic diversity and phenotypes?

6. References

Afzal AJ, Natarajan A, Saini N, Iqbal MJ, Geisler M, El Shemy HA, Mungur R, Willmitzer L, Lightfoot DA (2009) The Nematode Resistance Allele at the rhg1 Locus Alters the Proteome and Primary Metabolism of Soybean Roots. Plant Physiology **151:** 1264-1280

Aitman TJ, Dong R, Vyse TJ, Norsworthy PJ, Johnson MD, Smith J, Mangion J, Roberton-Lowe C, Marshall AJ, Petretto E, Hodges MD, Bhangal G, Patel SG, Sheehan-Rooney K, Duda M, Cook PR, Evans DJ, Domin J, Flint J, Boyle JJ, Pusey CD, Cook HT (2006) Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. Nature **439:** 851-855

Alkan C, Sajjadian S, Eichler EE (2011) Limitations of next-generation genome sequence assembly. Nat Methods **8:** 61-65

Alkharouf NW, Klink VP, Chouikha IB, Beard HS, MacDonald MH, Meyer S, Knap HT, Khan R, Matthews BF (2006) Timecourse microarray analyses reveal global changes in gene expression of susceptible Glycine max (soybean) roots during infection by Heterodera glycines (soybean cyst nematode). Planta **224:** 838-852

Alvarez JP, Pekker I, Goldshmidt A, Blum E, Amsellem Z, Eshed Y (2006) Endogenous and synthetic microRNAs stimulate simultaneous, efficient, and localized regulation of multiple targets in diverse species. Plant Cell **18:** 1134-1151

Anand SC (1992) REGISTRATION OF HARTWIG SOYBEAN. Crop Science **32:** 1069-1070

Arelli APR, Webb DM (1996) Molecular genetic diversity among soybean plant introductions with resistance to Heterodera glycines. Current Science **71:** 230-233

Arelli PR, Sleper DA, Yue P, Wilcox JA (2000) Soybean reaction to Races 1 and 2 of Heterodera glycines. Crop Science **40:** 824-826

Barnard RJ, Morgan A, Burgoyne RD (1996) Domains of alpha-SNAP required for the stimulation of exocytosis and for N-ethylmalemide-sensitive fusion protein (NSF) binding and activation. Mol Biol Cell **7:** 693-701

Barnard RJ, Morgan A, Burgoyne RD (1997) Stimulation of NSF ATPase activity by alpha-SNAP is required for SNARE complex disassembly and exocytosis. J Cell Biol **139:** 875-883

Baulcombe D (2004) RNA silencing in plants. Nature **431:** 356-363

Baulcombe D (2005) RNA silencing. Trends Biochem Sci **30:** 290-293

Baum TJ, Wubben MJE, Hardy KA, Su H, Rodermel SR (2000) A screen for Arabidopsis thaliana mutants with altered susceptibility to Heterodera schachtii. Journal of Nematology **32:** 166-173

Bent AF (1996) Plant disease resistance genes: function meets structure. Plant Cell **8:** 1757-1771

Boisson-Dernier A, Chabaud M, Garcia F, Becard G, Rosenberg C, Barker DG (2001) Agrobacterium rhizogenes-transformed roots of Medicago truncatula for the study of nitrogen-fixing and endomycorrhizal symbiotic associations. Molecular Plant-Microbe Interactions **14:** 695-700

Brown CJ, Todd KM, Rosenzweig RF (1998) Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment. Molecular biology and evolution **15:** 931-942

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

Brucker E, Niblack T, Kopisch-Obuch FJ, Diers BW (2005) The effect of rhg1 on reproduction of Heterodera glycines in the field and greenhouse and associated effects on agronomic traits. Crop Science **45:** 1721-1727

Byrd DW, Kirkpatrick T, Barker KR (1983) An improved technique for clearing and staining plant-tissue for detection of nematodes. Journal of Nematology **15:** 142-143

Cai DG, Kleine M, Kifle S, Harloff HJ, Sandal NN, Marcker KA, KleinLankhorst RM, Salentijn EMJ, Lange W, Stiekema WJ, Wyss U, Grundler FMW, Jung C (1997) Positional cloning of a gene for nematode resistance in sugar beet. Science **275:** 832-834

Caldwell BE, Brim CA, Ross JP (1960) Inheritance of resistance of soybeans to the cyst nematode, Heterodera glycines. Agronomy Journal **52:** 635-636

Calwell BE, Brim CA, Ross JP (1960) Inheritance of resistance of soybean to the cyst nematode, Heterodera glycines. Agronomy Journal **52:** 635-636

Cary T, Diers B (2013) 2012 Northern regional soybean cyst nematode tests. *In*. University of Illinios, Urbana, IL

Cary T, Diers BW (2010) 2009 Northern regional soybean cyst nematode tests. *In*. University of Illinios, Urbana, IL

Cary T, Diers BW (2011) 2010 Northern regional soybean cyst nematode tests. *In*. University of Illinios, Urbana, IL

Cary T, Diers BW (2012) 2011 Northern regional soybean cyst nematode tests. *In*. University of Illinios, Urbana, IL

Chang AYF, Liao BY (2012) DNA Methylation Rebalances Gene Dosage after Mammalian Gene Duplications. Molecular biology and evolution **29:** 133-144

Chen ZZ, Cheng CHC, Zhang JF, Cao LX, Chen L, Zhou LH, Jin YD, Ye H, Deng C, Dai ZH, Xu QH, Hu P, Sun SH, Shen Y, Chen LB (2008) Transcrintomic and genomic evolution under constant cold in Antarctic notothenioid fish. Proceedings of the National Academy of Sciences of the United States of America **105:** 12944-12949

Cheng Z, Buell CR, Wing RA, Jiang J (2002) Resolution of fluorescence in-situ hybridization mapping on rice mitotic prometaphase chromosomes, meiotic pachytene chromosomes and extended DNA fibers. Chromosome Res **10:** 379-387

Chia JM, Song C, Bradbury PJ, Costich D, de Leon N, Doebley J, Elshire RJ, Gaut B, Geller L, Glaubitz JC, Gore M, Guill KE, Holland J, Hufford MB, Lai JS, Li M, Liu X, Lu YL, McCombie R, Nelson R, Poland J, Prasanna BM, Pyhajarvi T, Rong TZ, Sekhon RS, Sun Q, Tenaillon MI, Tian F, Wang J, Xu X, Zhang ZW, Kaeppler SM, Ross-Ibarra J, McMullen MD, Buckler ES, Zhang GY, Xu YB, Ware D (2012) Maize HapMap2 identifies extant variation from a genome in flux. Nature Genetics **44:** 803-U238

Chiera JM, Bouchard RA, Dorsey SL, Park E, Buenrostro-Nava MT, Ling PP, Finer JJ (2007) Isolation of two highly active soybean (Glycine max (L.) Merr.) promoters and their characterization using a new automated image collection and analysis system. Plant Cell Reports **26:** 1501-1509

Cho HJ, Farrand SK, Noel GR, Widholm JM (2000) High-efficiency induction of soybean hairy roots and propagation of the soybean cyst nematode. Planta **210:** 195-204

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu XY, Ruden DM (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w(1118); iso-2; iso-3. Fly **6:** 80-92

Cloud GL, Riggs RD, Caviness CE (1988) Variability in host preference among field populations of Heterodera glycines. Journal of Nematology **20:** 417-420

Colgrove AL, Niblack TL (2008) Correlation of female indices from virulence assays on inbred lines and field Populations of Heterodera glycines. Journal of Nematology **40:** 39-45

Conant GC, Wolfe KH (2008) Turning a hobby into a job: How duplicated genes find new functions. Nature Reviews Genetics **9:** 938-950

Concibido VC, Diers BW, Arelli PR (2004) A decade of QTL mapping for cyst nematode resistance in soybean. Crop Science **44:** 1121-1131

Concibido VC, Lange DA, Denny RL, Orf JH, Young ND (1997) Genome mapping of soybean cyst nematode resistance genes in Peking, PI 90763, and PI 88788 using DNA markers. Crop Sci. **37:** 258-264

Conrad B, Antonarakis SE (2007) Gene duplication: A drive for phenotypic diversity and cause of human disease. *In* Annual Review of Genomics and Human Genetics, Vol 8. Annual Reviews, Palo Alto, pp 17-35

Constantin GD, Gronlund M, Johansen IE, Stougaard J, Lund OS (2008) Virus-induced gene silencing (VIGS) as a reverse genetic tool to study development of symbiotic root nodules. Molecular Plant-Microbe Interactions **21:** 720-727

Cook DE, Lee TG, Guo X, Melito S, Wang K, Bayless AM, Wang J, Hughes TJ, Willis DK, Clemente TE, Diers BW, Jiang J, Hudson ME, Bent AF (2012)

Copy number variation of multiple genes at Rhg1 mediates nematode resistance in soybean. Science **338:** 1206-1209

Craven SH, Neidle EL (2007) Double trouble: medical implications of genetic duplication and amplification in bacteria. Future Microbiology **2:** 309-321

Cregan P, Diers B (unpublished) Whole genome sequence analysis of SoyNAM: soybean nested association mapping population. Unpublished

Cregan PB, Mudge J, Fickus EW, Danesh D, Denny R, Young ND (1999) Two simple sequence repeat markers to select for soybean cyst nematode resistance coditioned by the rhg1 locus. Theoretical and Applied Genetics **99:** 811-818

Dassanayake M, Oh DH, Haas JS, Hernandez A, Hong H, Ali S, Yun DJ, Bressan RA, Zhu JK, Bohnert HJ, Cheeseman JM (2011) The genome of the extremophile crucifer Thellungiella parvula. Nature Genetics **43:** 913-U137

Davis EL, Hussey RS, Baum TJ (2004) Getting to the roots of parasitism by nematodes. Trends in Parasitology **20:** 134-141

Davis EL, Hussey RS, Mitchum MG, Baum TJ (2008) Parasitism proteins in nematode-plant interactions. Current Opinion in Plant Biology **11:** 360-366

DellaPenna D, O'Connor SE (2012) Plant science. Plant gene clusters and opiates. Science **336:** 1648-1649

Demuth JP, Hahn MW (2009) The life and death of gene families. Bioessays **31:** 29-39

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics **43:** 491-+

Diers BW, Cary T, Thomas D, Colgrove A, Niblack T (2010) Registration of LD00-2817P Soybean Germplasm Line with Resistance to Soybean Cyst Nematode from PI 437654. Journal of Plant Registrations **4:** 141-144

Diers BW, Skorupska HT, RaoArelli AP, Cianzio SR (1997) Genetic relationships among soybean plant introductions with resistance to soybean cyst nematodes. Crop Science **37:** 1966-1972

Dodds PN, Rathjen JP (2010) Plant immunity: towards an integrated view of plant-pathogen interactions. Nat Rev Genet **11:** 539-548

Dowen RH, Pelizzola M, Schmitz RJ, Lister R, Dowen JM, Nery JR, Dixon JE, Ecker JR (2012) Widespread dynamic DNA methylation in response to biotic stress. Proceedings of the National Academy of Sciences of the United States of America **109:** E2183-E2191

Doyle JJ, Dickson EE (1987) Preservation of plant-samples for DNA restriction endonuclease analysis. Taxon **36:** 715-722

Doyle JJ, Doyle JL, Harbison C (2003) Chloroplast-expressed glutamine synthetase in Glycine and related Leguminosae: Phylogeny, gene duplication, and ancient polyploidy. Systematic Botany **28:** 567-577

Doyle JJ, Luckow MA (2003) The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. Plant Physiology **131:** 900-910

Ellis J, Dodds P, Pryor T (2000) Structure, function and evolution of plant disease resistance genes. Curr Opin Plant Biol **3:** 278-284

Endo BY (1965) HISTOLOGICAL RESPONSES OF RESISTANT AND SUSCEPTIBLE SOYBEAN VARIETIES AND BACKCROSS PROGENY TO ENTRY AND DEVELOPMENT OF HETERODERA GLYCINES. Phytopathology **55:** 375-&

Endo BY (1978) FEEDING PLUG FORMATION IN SOYBEAN ROOTS INFECTED WITH THE SOYBEAN CYST NEMATODE. Phytopathology **68:** 1022-1031

Endo BY (1984) Ultrastructure of the esophagus of larvae of the soybean cyst nematode, Heterodera glycines. Proceedings of the Helminthological Society of Washington **51:** 1-24

Ernst K, Kumar A, Kriseleit D, Kloos DU, Phillips MS, Ganal MW (2002) The broad-spectrum potato cyst nematode resistance gene (Hero) from tomato is the only member of a large gene family of NBS-LRR genes with an unusual amino acid repeat in the LRR region. Plant Journal **31:** 127-136

Feuk L, Carson AR, Scherer SW (2006) Structural variation in the human genome. Nature Reviews Genetics **7:** 85-97

Finkers-Tomczak A, Bakker E, de Boer J, van der Vossen E, Achenbach U, Golas T, Suryaningrat S, Smant G, Bakker J, Goverse A (2011) Comparative sequence analysis of the potato cyst nematode resistance locus H1 reveals a major lack of co-linearity between three haplotypes in potato (Solanum tuberosum ssp.). Theoretical and Applied Genetics **122:** 595-608

Flagel LE, Wendel JF (2009) Gene duplication and evolutionary novelty in plants. New Phytologist **183:** 557-564

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

Frey M, Chomet P, Glawischnig E, Stettner C, Grun S, Winklmair A, Eisenreich W, Bacher A, Meeley RB, Briggs SP, Simcox K, Gierl A (1997) Analysis of a chemical plant defense mechanism in grasses. Science **277:** 696-699

Gheysen G, Mitchum MG (2011) How nematodes manipulate plant development pathways for infection. Current Opinion in Plant Biology **14:** 415-421

Girirajan S, Campbell CD, Eichler EE (2011) Human Copy Number Variation and Complex Genetic Disease. Annual Review Genetics, Vol 45 **45:** 203-226

Glover KD, Wang D, Arelli PR, Carlson SR, Cianzio SR, Diers BW (2004) Near isogenic lines confirm a soybean cyst nematode resistance gene from PI 88788 on linkage group J. Crop Science **44:** 936-941

Graham TL, Graham MY, Subramanian S, Yu O (2007) RNAi silencing of genes for elicitation or biosynthesis of 5-deoxyisoflavonoids suppresses race-specific resistance and hypersensitive cell death in Pytophthora sojae infected tissues. Plant Physiology **144:** 728-740

GRIN Germplasm reseources infomation network. *In*,

Guo B, Sleper DA, Lu P, Shannon JG, Nguyen HT, Arelli PR (2006) QTLs associated with resistance to soybean cyst nematode in soybean: Meta-analysis of QTL locations. Crop Science **46:** 595-602

Halbrendt JM, Lewis SA, Shipe ER (1992) A TECHNIQUE FOR EVALUATING HETERODERA-GLYCINES DEVELOPMENT IN SUSCEPTIBLE AND RESISTANT SOYBEAN. Journal of Nematology **24:** 84-91

Hammond-Kosack KE, Jones JDG (1997) Plant disease resistance genes. Annu. Rev. Plant Mol. Biol. **48:** 575-607

Hanson AD, Roje S (2001) One-carbon metabolism in higher plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. **52:** 119-137

Hastings PJ, Lupski JR, Rosenberg SM, Ira G (2009) Mechanisms of change in gene copy number. Nature Reviews Genetics **10:** 551-564

Hauge BM, M.L. W, Parsons JD, Parnell LD (2001) Nucleic acid molecules and other molecules associated with soybean cyst nematode resistance. *In*. Monsanto Company, USA

Hauge BM, Wang ML, Parsons JD, Parnell L (2006) Nucleaic acid molecules and other molecules associated with soybean cyst nematode resistance. . *In*,

Hebsgaard SM, Korning PG, Tolstrup N, Engelbrecht J, Rouze P, Brunak S (1996) Splice site prediction in Arabidopsis thaliana pre-mRNA by combining local and global sequence information. Nucleic Acids Research **24:** 3439-3452

Hernandez-Garcia CM, Martinelli AP, Bouchard RA, Finer JJ (2009) A soybean (Glycine max) polyubiquitin promoter gives strong constitutive expression in transgenic soybean. Plant Cell Reports **28:** 837-849

Hofgen R, Willmitzer L (1988) Storage of competent cells for Agrobacterium transformation. Nucleic Acids Research **16:** 9877-9877

Hu RB, Fan CM, Li HY, Zhang QZ, Fu YF (2009) Evaluation of putative reference genes for gene expression normalization in soybean by quantitative real-time RT-PCR. Bmc Molecular Biology **10:** 12

Innan H, Kondrashov F (2010) The evolution of gene duplications: classifying and distinguishing between models. Nature Reviews Genetics **11:** 97-108

Innes RW, Ameline-Torregrosa C, Ashfield T, Cannon E, Cannon SB, Chacko B, Chen NWG, Couloux A, Dalwani A, Denny R, Deshpande S, Egan AN, Glover N, Hans CS, Howell S, Ilut D, Jackson S, Lai HS, Mammadov J, del Campo SM, Metcalf M, Nguyen A, O'Bleness M, Pfeil BE, Podicheti R, Ratnaparkhe MB, Samain S, Sanders I, Segurens B, Sevignac M, Sherman-Broyles S, Thareau V, Tucker DM, Walling J, Wawrzynski A, Yi J, Doyle JJ, Geffroy V, Roe BA, Maroof MAS, Young ND (2008) Differential Accumulation of Retroelements and Diversification of NB-LRR Disease Resistance Genes in Duplicated Regions following Polyploidy in the Ancestor of Soybean. Plant Physiology **148:** 1740-1759

Iskow RC, Gokcumen O, Lee C (2012) Exploring the role of copy number variants in human adaptation. Trends in Genetics **28:** 245-257

Ithal N, Recknor J, Nettleton D, Maier T, Baum TJ, Mitchum MG (2007) Developmental transcript profiling of cyst nematode feeding cells in soybean roots. Molecular Plant-Microbe Interactions **20:** 510-525

Ivashuta S, Liu JY, Liu JQ, Lohar DP, Haridas S, Bucciarelli B, VandenBosch KA, Vance CP, Harrison MJ, Gantt JS (2005) RNA interference identifies a calcium-dependent protein kinase involved in Medicago truncatula root development. Plant Cell **17:** 2911-2921

Jackson SA, Wang ML, Goodman HM, Jiang J (1998) Application of fiber-FISH in physical mapping of Arabidopsis thaliana. Genome **41:** 566-572

Jahn R, Scheller RH (2006) SNAREs - engines for membrane fusion. Nature Reviews Molecular Cell Biology **7:** 631-643

Kandoth PK, Ithal N, Recknor J, Maier T, Nettleton D, Baum TJ, Mitchum MG (2011) The Soybean Rhg1 Locus for Resistance to the Soybean Cyst Nematode Heterodera glycines Regulates the Expression of a Large Number of Stress- and Defense-Related Genes in Degenerating Feeding Cells. Plant Physiology **155:** 1960-1975

Kelley LA, Sternberg MJE (2009) Protein structure prediction on the Web: a case study using the Phyre server. Nature Protocols **4:** 363-371

Kenrick P, Crane PR (1997) The origin and early evolution of plants on land. Nature **389:** 33-39

Kereszt A, Li DX, Indrasumunar A, Nguyen CDT, Nontachaiyapoom S, Kinkema M, Gresshoff PM (2007) Agrobacterium rhizogenes - mediated transformation of soybean to study root biology. Nat Protoc **2:** 948-952

Khraiwesh B, Ossowski S, Weigel D, Reski R, Frank W (2008) Specific gene silencing by artificial microRNAs in Physcomitrella patens: An alternative to targeted gene knockouts. Plant Physiology **148:** 684-693

Kim DG, Riggs RD, Mauromoustakos A (1998) Variation in resistance of soybean lines to races of Heterodera glycines. Journal of Nematology **30:** 184-191

Kim M, Hyten DL, Niblack TL, Diers BW (2011) Stacking Resistance Alleles from Wild and Domestic Soybean Sources Improves Soybean Cyst Nematode Resistance. Crop Science **51:** 934-943

Kim M-S, Hyten DL, Bent AF, Diers BW (2010) Fine mapping of the SCN resistance locus rhg1-b from PI 88788. Plant Genome **3:** 81-89

Kim MY, Lee S, Van K, Kim TH, Jeong SC, Choi IY, Kim DS, Lee YS, Park D, Ma J, Kim WY, Kim BC, Park S, Lee KA, Kim DH, Kim KH, Shin JH, Jang YE, Do Kim K, Liu WX, Chaisan T, Kang YJ, Lee YH, Moon JK, Schmutz J, Jackson SA, Bhak J, Lee SH (2010) Whole-genome sequencing and intensive analysis of the undomesticated soybean (Glycine soja Sieb. and Zucc.) genome. Proceedings of the National Academy of Sciences of the United States of America **107:** 22032-22037

Kim YH, Kim KS, Riggs RD (2010) Differential Subcellular Responses in Resistance Soybeans Infected with Soybean Cyst Nematode Races. Plant Pathology Journal **26:** 154-158

Kim YH, Riggs RD, Kim KS (1987) Structural changes associated with resistance of soybean to Heterodera glycines. Journal of Nematology **19:** 177-187

Klink VP, Hosseini P, Matsye P, Alkharouf NW, Matthews BF (2009) A gene expression analysis of syncytia laser microdissected from the roots of the Glycine max (soybean) genotype PI 548402 (Peking) undergoing a resistant reaction after infection by Heterodera glycines (soybean cyst nematode). Plant Molecular Biology **71:** 525-567

Klink VP, Hosseini P, Matsye PD, Alkharouf NW, Matthews BF (2010) Syncytium gene expression in Glycine max([PI 88788]) roots undergoing a resistant reaction to the parasitic nematode Heterodera glycines. Plant Physiology and Biochemistry **48:** 176-193

Klink VP, Hosseini P, Matsye PD, Alkharouf NW, Matthews BF (2011) Differences in gene expression amplitude overlie a conserved transcriptomic program occurring between the rapid and potent localized resistant reaction at the syncytium of the Glycine max genotype Peking (PI 548402) as compared to the prolonged and potent resistant reaction of PI 88788. Plant Molecular Biology **75:** 141-165

Klink VP, Overall CC, Alkharouf NW, MacDonald MH, Matthews BF (2007) A time-course comparative microarray analysis of an incompatible and compatible response by Glycine max (soybean) to Heterodera glycines (soybean cyst nematode) infection. Planta **226:** 1423-1447

Klink VP, Overall CC, Alkharouf NW, MacDonald MH, Matthews BF (2007) Laser capture microdissection (LCM) and comparative microarray expression analysis of syncytial cells isolated from incompatible and compatible soybean (Glycine max) roots infected by the soybean cyst nematode (Heterodera glycines). Planta **226:** 1389-1409

Kondrashov FA (2012) Gene duplication as a mechanism of genomic adaptation to a changing environment. Proceedings of the Royal Society B-Biological Sciences **279:** 5048-5057

Kondrashov FA, Kondrashov AS (2006) Role of selection in fixation of gene duplications. Journal of Theoretical Biology **239:** 141-151

Kondrashov FA, Rogozin IB, Wolf YI, Koonin EV (2002) Selection in the evolution of gene duplications. Genome Biol **3:** 9

Kulkarni SS, Bajwa SG, Robbins RT, Costello TA, Kirkpatrick TL (2008) Effect of Soybean Cyst Nematode (Heterodera Glycines) Resistance Rotation on Scn Population Distribution, Soybean Canopy Reflectance, and Grain Yield. Transactions of the Asabe **51:** 1511-1517

Labbe P, Berticat C, Berthomieu A, Unal S, Bernard C, Weill M, Lenormand T (2007) Forty years of erratic insecticide resistance evolution in the Mosquito Culex pipiens. Plos Genetics **3:** 2190-2199

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods **9:** 357-U354

Lauritis JA, Rebois RV, Graney LS (1983) Development of Heterodera glycines ichinohe on soybean, Glycine max (L) merr, under gnotobiotic conditions Journal of Nematology **15:** 272-281

Lauritis JA, Rebois RV, Graney LS (1983) Development of Heterodera glycines ichinohe on soybean, Glycine max (L) merr, under gnotobiotic conditions. . Journal of Nematology **15:** 272-281

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25:** 1754-1760

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics **25:** 2078-2079

Li YH, Chen SY, Young ND (2004) Effect of the rhg1 gene on penetration, development and reproduction of Heterodera glycines race 3. Nematology **6:** 729-736

Libault M, Farmer A, Brechenmacher L, Drnevich J, Langley RJ, Bilgin DD, Radwan O, Neece DJ, Clough SJ, May GD, Stacey G (2010) Complete Transcriptome of the Soybean Root Hair Cell, a Single-Cell Model, and Its Alteration in Response to Bradyrhizobium japonicum Infection. Plant Physiology **152:** 541-552

Lightfoot DA, Meksem K (2002) Isolated polynucleotides and polypeptides relating to loci underlying resistance to soybean cyst nematode and soybean sudden death syndrome and methods employing same. *In*. Southern Illinois University, USA

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

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science **290:** 1151-1155

M.M. (2012) Materials and methods are available as supplementary materials on Science Online.

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

Mankin SL, Thompson WF (2001) New green fluorescent protein genes for plant transformation: Intron-containing, ER-localized, and soluble-modified. Plant Mol. Bio. Rep. **19:** 13-26

Maron LG, Guimaraes CT, Kirst M, Albert PS, Birchler JA, Bradbury PJ, Buckler ES, Coluccio AE, Danilova TV, Kudrna D, Magalhaes JV, Pineros MA, Schatz MC, Wing RA, Kochian LV (2013) Aluminum tolerance in maize is associated with higher MATE1 gene copy number. Proceedings of the

National Academy of Sciences of the United States of America **110:** 5241-5246

Marques-Bonet T, Girirajan S, Eichler EE (2009) The origins and impact of primate segmental duplications. Trends in Genetics **25:** 443-454

Matson AL, Williams LF (1965) Evidence of a fourth gene for resistance to the soybean cyst nematode. Crop Sci. **5:** 477

Matsye PD, Kumar R, Hosseini P, Jones CM, Tremblay A, Alkharouf NW, Matthews BF, Klink VP (2011) Mapping cell fate decisions that occur during soybean defense responses. Plant Molecular Biology **77:** 513-528

Matsye PD, Lawrence GW, Youssef RM, Kim KH, Lawrence KS, Matthews BF, Klink VP (2012) The expression of a naturally occurring, truncated allele of an alpha-SNAP gene suppresses plant parasitic nematode infection. Plant Mol Biol **80:** 131-155

Mazarei M, Miao B, Paiva NL (1996) Co-transormed hairy roots of soybean: a rapid system for in vivo plant promoter analysis. Phytpathology **86**

McHale LK, Haun WJ, Xu WW, Bhaskar PB, Anderson JE, Hyten DL, Gerhardt DJ, Jeddeloh JA, Stupar RM (2012) Structural variants in the soybean genome localize to clusters of biotic stress response genes. Plant Physiol

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research **20:** 1297-1303

Meksem K, Pantazopoulos P, Njiti VN, Hyten LD, Arelli PR, Lightfoot DA (2001) 'Forrest' resistance to the soybean cyst nematode is bigenic: saturation mapping of the Rhg1 and Rhg4 loci. Theoretical and Applied Genetics **103:** 710-717

Melito S, Heuberger AL, Cook D, Diers BW, MacGuidwin AE, Bent AF (2010) A nematode demographics assay in transgenic roots reveals no significant impacts of the Rhg1 locus LRR-Kinase on soybean cyst nematode resistance. Bmc Plant Biology **10**

Mikel MA, Diers BW, Nelson RL, Smith HH (2010) Genetic Diversity and Agronomic Improvement of North American Soybean Germplasm. Crop Science **50:** 1219-1229

Mitchum MG, Wang XH, Wang JY, Davis EL (2012) Role of Nematode Peptides and Other Small Molecules in Plant Parasitism. *In* Annual Review of Phytopathology, Vol 50, Vol 50, pp 175-195

Moore RC, Purugganan MD (2005) The evolutionary dynamics of plant duplicate genes. Current Opinion in Plant Biology **8:** 122-128

Morgan A, Dimaline R, Burgoyne RD (1994) The ATPase activity of N-ethylmaleimide-sensitive fusion protein (NSF) is regulated by soluble NSF attachment proteins. J Biol Chem **269:** 29347-29350

Mrosk C, Forner S, Hause G, Kuster H, Kopka J, Hause B (2009) Composite Medicago truncatula plants harbouring Agrobacterium rhizogenes-

transformed roots reveal normal mycorrhization by Glomus intraradices. Journal of Experimental Botany **60:** 3797-3807

Murray MG, Thompson WF (1980) RAPID ISOLATION OF HIGH MOLECULAR-WEIGHT PLANT DNA. Nucleic Acids Research **8:** 4321-4325

Nagamatsu A, Masuta C, Matsuura H, Kitamura K, Abe J, Kanazawa A (2009) Down-regulation of flavonoid 3 '-hydroxylase gene expression by virus-induced gene silencing in soybean reveals the presence of a threshold mRNA level associated with pigmentation in pubescence. J Plant Physiol **166:** 32-39

Nagamatsu A, Masuta C, Senda M, Matsuura H, Kasai A, Hong S, Kitamura K, Abe J, Kanazawa A (2007) Functional analysis of soybean genes involved in flavonoid biosynthesis by virus-induced gene silencing. Plant Biotechnology Journal **5:** 778-790

Nakagawa T, Kurose T, Hino T, Tanaka K, Kawamukai M, Niwa Y, Toyooka K, Matsuoka K, Jinbo T, Kimura T (2007) Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. J Biosci Bioeng **104:** 34-41

Narayanan R, Atz R, Denny R, Young N, Somers D (1999) Expression of soybean cyst nematode resistance in transgenic hairy roots of soybean. Crop Science **39:** 1680-1686

NASS (2012) Acreage *In* NAS Service, ed,

Niblack TL (2005) Soybean cyst nematode management reconsidered. Plant Disease **89:** 1020-1026

Niblack TL, Arelli PR, Noel GR, Opperman CH, Ore JH, Schmitt DP, Shannon JG, Tylka GL (2002) A revised classification scheme for genetically diverse populations of Heterodera glycines. Journal of Nematology **34:** 279-288

Niblack TL, Colgrove AL, Colgrove K, Bond JP (2008) Shift in virulance of soybean cyst nematode is associated with use of resistance from PI 88788. Plant Health Progress

Niblack TL, Lambert KN, Tylka GL (2006) A model plant pathogen from the kingdom animalia: Heterodera glycines, the soybean cyst nematode. Annual Review of Phytopathology **44:** 283-303

Niblack TL, Tylka GL, Riggs RD (2004) Nematode Pathogens of Soybean. *In* HR Boerma, JE Specht, eds, Soybeans: Improvement, Production and Uses, Ed Third Edition. American Society of Agronomy, Madison, WI, pp 821-852

Oh DH, Dassanayake M, Bohnert HJ, Cheeseman JM (2012) Life at the extreme: lessons from the genome. Genome Biol **13**

Ohno S (1970) Evolution by gene duplication. Berlin, Germany: Springer

Okumoto S, Pilot G (2011) Amino Acid Export in Plants: A Missing Link in Nitrogen Cycling. Molecular Plant **4:** 453-463

Olhoft PM, Bernal LM, Grist LB, Hill DS, Mankin SL, Shen YW, Kalogerakis M, Wiley H, Toren E, Song HS, Hillebrand H, Jones T (2007) A novel Agrobacterium rhizogenes-mediated transformation method of soybean

[Glycine max (L.) Merrill] using primary-node explants from seedlings. In Vitro Cellular & Developmental Biology-Plant **43:** 536-549

Ossowski S, Schwab R, Weigel D (2008) Gene silencing in plants using artificial microRNAs and other small RNAs. Plant J **53:** 674-690

Paal J, Henselewski H, Muth J, Meksem K, Menendez CM, Salamini F, Ballvora A, Gebhardt C (2004) Molecular cloning of the potato Gro1-4 gene conferring resistance to pathotype Ro1 of the root cyst nematode Globodera rostochiensis, based on a candidate gene approach. Plant Journal **38:** 285-297

Pearce S, Saville R, Vaughan SP, Chandler PM, Wilhelm EP, Sparks CA, Al-Kaff N, Korolev A, Boulton MI, Phillips AL, Hedden P, Nicholson P, Thomas SG (2011) Molecular Characterization of Rht-1 Dwarfing Genes in Hexaploid Wheat. Plant Physiology **157:** 1820-1831

Peiter E, Sun J, Heckmann AB, Venkateshwaran M, Riely BK, Otegui MS, Edwards A, Freshour G, Hahn MG, Cook DR, Sanders D, Oldroyd GED, Downie JA, Ane JM (2007) The Medicago truncatula DMI1 protein modulates cytosolic calcium signaling. Plant Physiology **145:** 192-203

Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, Werner J, Villanea FA, Mountain JL, Misra R, Carter NP, Lee C, Stone AC (2007) Diet and the evolution of human amylase gene copy number variation. Nature Genetics **39:** 1256-1260

Perry GH, Yang F, Marques-Bonet T, Murphy C, Fitzgerald T, Lee AS, Hyland C, Stone AC, Hurles ME, Tyler-Smith C, Eichler EE, Carter NP, Lee C, Redon R (2008) Copy number variation and evolution in humans and chimpanzees. Genome Research **18:** 1698-1710

Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP (2004) Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper - Excel-based tool using pair-wise correlations. Biotechnology Letters **26:** 509-515

Qian WF, Liao BY, Chang AYF, Zhang JZ (2010) Maintenance of duplicate genes and their functional redundancy by reduced expression. Trends in Genetics **26:** 425-430

Qiu BX, Arelli PR, Sleper DA (1999) RFLP markers associated with soybean cyst nematode resistance and seed composition in a 'Peking' x 'Essex' population. Theoretical and Applied Genetics **98:** 356-364

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics **26:** 841-842

Raoarelli AP, Wrather JA, Anand SC (1992) GENETIC DIVERSITY AMONG ISOLATES OF HETERODERA-GLYCINES AND SOURCES OF RESISTANCE IN SOYBEANS. Plant Disease **76:** 894-896

Rice LM, Brunger AT (1999) Crystal structure of the vesicular transport protein Sec17: implications for SNAP function in SNARE complex disassembly. Mol Cell **4:** 85-95

Riggs RD, Schmitt DP (1988) COMPLETE CHARACTERIZATION OF THE RACE SCHEME FOR HETERODERA-GLYCINES. Journal of Nematology **20:** 392-395

Ruben E, Jamai A, Afzal J, Njiti VN, Triwitayakorn K, Iqbal MJ, Yaegashi S, Bashir R, Kazi S, Arelli P, Town CD, Ishihara H, Meksem K, Lightfoot DA (2006) Genomic analysis of the rhg1 locus: candidate genes that underlie soybean resistance to the cyst nematode. Molecular Genetics and Genomics **276:** 503-516

Sambrook J, Russell DW (2001) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, New York

Schlattl A, Anders S, Waszak SM, Huber W, Korbel JO (2011) Relating CNVs to transcriptome data at fine resolution: Assessment of the effect of variant size, type, and overlap with functional regions. Genome Research **21:** 2004-2013

Schmidt JM, Good RT, Appleton B, Sherrard J, Raymant GC, Bogwitz MR, Martin J, Daborn PJ, Goddard ME, Batterham P, Robin C (2010) Copy Number Variation and Transposable Elements Feature in Recent, Ongoing Adaptation at the Cyp6g1 Locus. Plos Genetics **6:** 11

Schmitt DP, Shannon G (1992) DIFFERENTIATING SOYBEAN RESPONSES TO HETERODERA GLYCINES RACES. Crop Science **32:** 275-277

Schmitz RJ, He Y, Valdez-Lopez O, Khan SM, Joshi T, Urich MA, Nery JR, Diers B, Xu D, Stacey G, Ecker J (2013) Epigenome-wide inheritance of cytosine methylation variants in a recombinant inbred population Genome Research

Schmutz J, Cannon SB, Schlueter J, Ma JX, Mitros T, Nelson W, Hyten DL, Song QJ, Thelen JJ, Cheng JL, Xu D, Hellsten U, May GD, Yu YS, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu SQ, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du JC, Tian ZX, Zhu LC, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the palaeopolyploid soybean (vol 463, pg 178, 2010). Nature **465:** 120-120

Schroeder NE, MacGuidwin AE (2006) Uptake of FITC by Heterodera glycines is inhibited by stomal occlusion and quiescence. Journal of Nematology **38:** 291-292

Schwab R, Ossowski S, Riester M, Warthmann N, Weigel D (2006) Highly specific gene silencing by artificial microRNAs in Arabidopsis. Plant Cell **18:** 1121-1133

Schwab R, Palatnik JF, Riester M, Schommer C, Schmid M, Weigel D (2005) Specific effects of MicroRNAs on the plant transcriptome. Developmental Cell **8:** 517-527

Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi MY, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M (2004) Large-

scale copy number polymorphism in the human genome. Science **305:** 525-528

Segovia M (1994) LEISHMANIA GENE AMPLIFICATION - A MECHANISM OF DRUG-RESISTANCE. Annals of Tropical Medicine and Parasitology **88:** 123-130

Severin AJ, Woody JL, Bolon YT, Joseph B, Diers BW, Farmer AD, Muehlbauer GJ, Nelson RT, Grant D, Specht JE, Graham MA, Cannon SB, May GD, Vance CP, Shoemaker RC (2010) RNA-Seq Atlas of Glycine max: A guide to the soybean transcriptome. BMC Plant Biology **10**

Sharma SB (1998) The Cyst Nematodes. Kluwer Acheademic Publishers

Shier M (2008) Soybean varieties with soybean cyst nematode resistance.

Sindhu AS, Maier TR, Mitchum MG, Hussey RS, Davis EL, Baum TJ (2009) Effective and specific in planta RNAi in cyst nematodes: expression interference of four parasitism genes reduces parasitic success. Journal of Experimental Botany **60:** 315-324

Small I (2007) RNAi for revealing and engineering plant gene functions. Current Opinion in Biotechnology **18:** 148-153

Smant G, Stokkermans J, Yan YT, de Boer JM, Baum TJ, Wang XH, Hussey RS, Gommers FJ, Henrissat B, Davis EL, Helder J, Schots A, Bakker J (1998) Endogenous cellulases in animals: Isolation of beta-1,4-endoglucanase genes from two species of plant-parasitic cyst nematodes. Proceedings of the National Academy of Sciences of the United States of America **95:** 4906-4911

Smith NA, Singh SP, Wang MB, Stoutjesdijk PA, Green AG, Waterhouse PM (2000) Total silencing by intron-spliced hairpin RNAs. Nature **407:** 319-320

Sobczak M, Avrova A, Jupowicz J, Phillips MS, Ernst K, Kumar A (2005) Characterization of susceptibility and resistance responses to potato cyst nematode (Globodera spp.) infection of tomato lines in the absence and presence of the broad-spectrum nematode resistance Hero gene. Molecular Plant-Microbe Interactions **18:** 158-168

Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, Specht JE, Cregan PB (2004) A new integrated genetic linkage map of the soybean. Theoretical and Applied Genetics **109:** 122-128

Springer NM, Ying K, Fu Y, Ji TM, Yeh CT, Jia Y, Wu W, Richmond T, Kitzman J, Rosenbaum H, Iniguez AL, Barbazuk WB, Jeddeloh JA, Nettleton D, Schnable PS (2009) Maize Inbreds Exhibit High Levels of Copy Number Variation (CNV) and Presence/Absence Variation (PAV) in Genome Content. Plos Genetics **5**

Stankiewicz P, Lupski JR (2010) Structural Variation in the Human Genome and its Role in Disease. Annual Review of Medicine **61:** 437-455

Steemans P, Le Herisse A, Melvin J, Miller MA, Paris F, Verniers J, Wellman CH (2009) Origin and Radiation of the Earliest Vascular Land Plants. Science **324:** 353-353

Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, Redon R, Bird CP, de Grassi A, Lee C, Tyler-Smith C, Carter N, Scherer SW, Tavare

S, Deloukas P, Hurles ME, Dermitzakis ET (2007) Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science **315:** 848-853

Subramanian S, Graham MY, Yu O, Graham TL (2005) RNA interference of soybean isoflavone synthase genes leads to silencing in tissues distal to the transformation site and to enhanced susceptibility to Phytophthora sojae. Plant Physiol **137:** 1345-1353

Swaminathan K, Varala K, Hudson ME (2007) Global repeat discovery and estimation of genomic copy number in a large, complex genome using a high-throughput 454 sequence survey. BMC genomics **8:** 13

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular biology and evolution **28:** 2731-2739

Triglia T, Foote SJ, Kemp DJ, Cowman AF (1991) IMPLICATION OF THE MULTIDRUG RESISTANCE GENE PFMDR1 IN PLASMODIUM-FALCIPARUM HAS ARISEN AS MULTIPLE INDEPENDENT EVENTS. Mol Cell Biol **11:** 5244-5250

Valdes-Lopez O, Thibivilliers S, Qiu J, Xu WWZ, Nguyen THN, Libault M, Le BH, Goldberg RB, Hill CB, Hartman GL, Diers B, Stacey G (2011) Identification of Quantitative Trait Loci Controlling Gene Expression during the Innate Immunity Response of Soybean. Plant Physiology **157:** 1975-1986

van der Vossen EAG, van der Voort J, Kanyuka K, Bendahmane A, Sandbrink H, Baulcombe DC, Bakker J, Stiekema WJ, Klein-Lankhorst RM (2000) Homologues of a single resistance-gene cluster in potato confer resistance to distinct pathogens: a virus and a nematode. Plant Journal **23:** 567-576

Walling JG, Jiang JM (2012) DNA and Chromatin Fiber-Based Plant Cytogenetics. *In* Plant Cytogenetics: Genome Structure and Chromosome Function, Vol 4. Springer, New York, pp 121-130

Warthmann N, Chen H, Ossowski S, Weigel D, Herve P (2008) Highly Specific Gene Silencing by Artificial miRNAs in Rice. PLoS One **3**

Wasson AP, Pellerone FI, Mathesius U (2006) Silencing the flavonoid pathway in Medicago truncatula inhibits root nodule formation and prevents auxin transport regulation by rhizobia. Plant Cell **18:** 1617-1629

Webb DM, Baltazar BM, Raoarelli AP, Schupp J, Clayton K, Keim P, Beavis WD (1995) Genetic-mapping of soybean cyst-nematode race-3 resistance loci in the soybean PI-437654. Theoretical and Applied Genetics **91:** 574-581

Wesley SV, Helliwell CA, Smith NA, Wang M, Rouse DT, Liu Q, Gooding PS, Singh SP, Abbott D, Stoutjesdijk PA, Robinson SP, Gleave AP, Green AG, Waterhouse PM (2001) Construct design for efficient, effective and high-throughput gene silencing in plants. Plant J **27:** 581-590.

Williamson VM, Hussey RS (1996) Nematode pathogenesis and resistance in plants. Plant Cell **8:** 1735-1745

Wingen LU, Munster T, Faigl W, Deleu W, Sommer H, Saedler H, Theissen G (2012) Molecular genetic basis of pod corn (Tunicate maize). Proceedings of the National Academy of Sciences of the United States of America **109:** 7115-7120

Winter SMJ, Shelp BJ, Anderson TR, Welacky TW, Rajcan I (2007) QTL associated with horizontal resistance to soybean cyst nematode in Glycine soja PI464925B. Theoretical and Applied Genetics **114:** 461-472

Wise AA, Liu Z, Binns AN (2006) Three methods for the introduction of foreign DNA into Agrobacterium. Methods Mol Biol **343:** 43-53

Wong ATS, Tylka GL, Hartzler RG (1993) Effects of 8 herbicides on in-vitro hatching of Heterodera-glycines. Journal of Nematology **25:** 578-584

Wrather JA, Koenning SR (2006) Estimates of disease effects on soybean yields in the United States 2003 to 2005. Journal of Nematology **38:** 173-180

Wrather JA, Koenning SR (2009) Effects of Diseases on Soybean Yields in the United States 1996 to 2007. Plant Health Progress

Wu XL, Blake S, Sleper DA, Shannon JG, Cregan P, Nguyen HT (2009) QTL, additive and epistatic effects for SCN resistance in PI 437654. Theoretical and Applied Genetics **118:** 1093-1105

Yen S, Chen P, Yen HCE (1999) Plant Gene Register PGR99-030. Cloning of a wound-induced gene WI12 from Mesembryanthemum crystallinum (accession no. AF117224). Plant Physiology **119:** 1147-1148

Young LD (1996) Yield loss in soybean caused by Heterodera glycines. Journal of Nematology **28:** 604-607

Yu J, Wang J, Lin W, Li SG, Li H, Zhou J, Ni PX, Dong W, Hu SN, Zeng CQ, Zhang JG, Zhang Y, Li RQ, Xu ZY, Li ST, Li XR, Zheng HK, Cong LJ, Lin L, Yin JN, Geng JN, Li GY, Shi JP, Liu J, Lv H, Li J, Deng YJ, Ran LH, Shi XL, Wang XY, Wu QF, Li CF, Ren XY, Wang JQ, Wang XL, Li DW, Liu DY, Zhang XW, Ji ZD, Zhao WM, Sun YQ, Zhang ZP, Bao JY, Han YJ, Dong LL, Ji J, Chen P, Wu SM, Liu JS, Xiao Y, Bu DB, Tan JL, Yang L, Ye C, Zhang JF, Xu JY, Zhou Y, Yu YP, Zhang B, Zhuang SL, Wei HB, Liu B, Lei M, Yu H, Li YZ, Xu H, Wei SL, He XM, Fang LJ, Zhang ZJ, Zhang YZ, Huang XG, Su ZX, Tong W, Li JH, Tong ZZ, Li SL, Ye J, Wang LS, Fang L, Lei TT, Chen C, Chen H, Xu Z, Li HH, Huang HY, Zhang F, Xu HY, Li N, Zhao CF, Dong LJ, Huang YQ, Li L, Xi Y, Qi QH, Li WJ, Hu W, Zhang YL, Tian XJ, Jiao YZ, Liang XH, Jin JA, Gao L, Zheng WM, Hao BL, Liu SQ, Wang W, Yuan LP, Cao ML, McDermott J, Samudrala R, Wong GKS, Yang HM (2005) The Genomes of Oryza sativa: A history of duplications. PLoS Biol **3:** 266-281

Yu P, Wang CH, Xu Q, Feng Y, Yuan XP, Yu HY, Wang YP, Tang SX, Wei XH (2011) Detection of copy number variations in rice using array-based comparative genomic hybridization. BMC genomics **12:** 8

Zhang W, He SY, Assmann SM (2008) The plant innate immunity response in stomatal guard cells invokes G-protein-dependent ion channel regulation. Plant J **56:** 984-996

Zhang ZH, Yu JY, Li DF, Zhang ZY, Liu FX, Zhou X, Wang T, Ling Y, Su Z (2010) PMRD: plant microRNA database. Nucleic Acids Research **38:** D806-D813

Zheng LY, Guo XS, He B, Sun LJ, Peng Y, Dong SS, Liu TF, Jiang SY, Ramachandran S, Liu CM, Jing HC (2011) Genome-wide patterns of genetic variation in sweet and grain sorghum (Sorghum bicolor). Genome Biol **12:** 14