

**Atypical Variants of the Soybean SNARE-(Soluble NSF Attachment Protein Receptors)-Recycling Machinery Underlie *Rhg1*-mediated Resistance to Soybean Cyst Nematode**

By

**Adam Milton Bayless**

A dissertation submitted in partial fulfillment of  
the requirements for the degree of

**Doctor of Philosophy**

**(Cellular and Molecular Biology)**

at the

**UNIVERSITY OF WISCONSIN-MADISON**

**2017**

Date of final oral examination: 08/24/2017

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

Andrew F. Bent, Professor, Plant Pathology  
Caitlyn Allen, Professor, Plant Pathology  
Jean-Michel Ané, Professor, Bacteriology  
Sebastian Y. Bednarek, Professor, Biochemistry  
Xuehua Zhong, Assistant Professor, Genetics

## Abstract

*Glycine max* (soybean) is the world's most widely grown legume. The most damaging soybean pathogen in terms of yield loss is the soybean cyst nematode (*Heterodera glycines*, SCN), which causes annual U.S. yield losses valuing over \$1 billion USD. *Glycine* species encode multiple defenses against SCN, but of these, the *Rhg1* (Resistance to Heterodera glycines 1) locus is the strongest known resistance locus, and is used in all commercial SCN-resistant soybeans. The *Rhg1* resistance phenotype triggers death of the nematode-induced feeding site, thereby shutting off nutrition from the now sedentary nematode. Previous studies misidentified *Rhg1* as a leucine-rich repeat receptor kinase, however, later genetic mapping studies excluded this kinase and defined *Rhg1* to a narrow genetic interval of 11 putative candidate genes. The focus of this dissertation has been to identify the gene(s) conferring *Rhg1* resistance and determine how they molecularly function during SCN-resistance.

We identified the genes conferring *Rhg1* function and discovered that *Rhg1* is an unusual disease resistance locus. *Rhg1* is an ~30 kb block of four different genes that is tandemly repeated up to 10 times. Gene silencing and resistance complementation indicated that three different *Rhg1* genes, including an usual  $\alpha$ -SNAP allele (alpha-Soluble-NSF Attachment Protein), contribute to SCN-resistance. Two distinct *Rhg1* phenotypic classes have long been known by soybean growers. In a follow up, we determined that *Rhg1* repeat copy number (*Rhg1* high copy: 4 or more blocks, *Rhg1* low copy: 3 or fewer blocks) defines these phenotypic *Rhg1* classes, and moreover, that *Rhg1* high copy vs. low copy haplotypes encode distinct polymorphic  $\alpha$ -SNAP proteins. Subsequently, we characterized both *Rhg1* high and low copy  $\alpha$ -SNAP proteins. We found that, unlike the WT *Rhg1*  $\alpha$ -SNAP protein encoded by SCN susceptible soybeans, either resistance-type  $\alpha$ -SNAP is impaired in normal interactions with the N-

ethylmaleimide Sensitive Factor (NSF). High expression of either *Rhg1* resistance-type  $\alpha$ -SNAP impeded WT NSF functions and hindered vesicular trafficking. Finally, within syncytia of the high copy *Rhg1* variety Fayette, we detected that the ratio of resistance-type  $\alpha$ -SNAPs increases relative to WT  $\alpha$ -SNAPs, suggesting that a semi-dominant negative mechanism dependent on  $\alpha$ -SNAP ratios may underlie *Rhg1* resistance. Later investigations with low copy *Rhg1* soybeans showed that overall WT  $\alpha$ -SNAP abundance was strikingly low compared to SCN-susceptible or high copy *Rhg1* soybeans. We then re-explored NSF loci in low copy *Rhg1* varieties and discovered a novel NSF allele with unique N-domain polymorphisms, which we termed RAN07 (*Rhg1* associated NSF on chromosome 07). We found that NSF<sub>RAN07</sub> was present within and needed for the viability of all soybean germplasm that carries high or low copy resistance-conferring *Rhg1*. Biochemical assays showed that NSF<sub>RAN07</sub> polymorphisms improve upon WT-NSF for compatibility with *Rhg1* resistance type  $\alpha$ -SNAPs. *In planta* studies revealed that NSF<sub>RAN07</sub> more effectively complements the cytotoxic properties of *Rhg1* resistance-type  $\alpha$ -SNAPs. We further demonstrated that a separate WT  $\alpha$ -SNAP locus on chromosome 11, which maps near a putative minor resistance QTL, does not produce a stable protein. These findings suggest that replacement and/or rewiring of both  $\alpha$ -SNAP and NSF – the core components of the SNARE recycling machinery – underlie *Rhg1*-mediated SCN resistance. Lastly, we discovered that low copy *Rhg1* haplotypes harbor an intact 4.77 kb Copia retrotransposon within intron 1 of the resistance type  $\alpha$ -SNAP. No significant impacts of this Copia element on low copy  $\alpha$ -SNAP production were observed. This *Rhg1*<sub>Low Copy</sub> associated Copia (RAC) was not detected in SCN-susceptible or *Rhg1* high copy soybeans, but was present in all three *Rhg1* repeats in all examined low copy *Rhg1* soybean accessions.

## Acknowledgments

I am incredibly grateful to my advisor Andrew Bent for the privilege to work in the lab. All research projects have ups and downs, but I feel especially fortunate to have been part of the Bent lab team investigating the endless surprises of *Rhg1*. Andrew has been an ideal mentor, full of patience, good advice and trust. Andrew's office was always open to bounce off ideas, get guidance or encouragement or to just chat. I especially want to thank my current committee members, Caitlyn Allen, Sebastian Bednarek, Jean-Michel Ané, and Xuehua Zhang, as well as past members, Amy Charkowski and Heidi Goodrich-Blair, for their expert advice, assistance and guidance during my graduate career.

To Bent lab members past and present: Junqi Song, David Cook, Teresa Koller, Yangrong Cao, Xiaoli Guo, Brian Kepper, Alice Teillet, Stephen Mosher, Katelyn Horgan, Patrick H. McMinn, Ryan Zapotocny, John Smith, Shaojie Han, Hu Die, and Derrick Grunwald - all of you have been wonderful to work with and I wish you all the very best (except of course for Derrick, "Haw-haw!" Derrick if you're reading this!). David and Xiaoli were especially helpful when I first joined the lab, and John Smith also contributed greatly to our work. Also, I want to give a big shout out to all the Bent lab undergrads - Kaela, Jackson, Amy, Jordyn, Anders, Joyce and Demi - for organizing and cleaning up our messes, and for just being really cool in general.

A tremendous thanks to UW-Madison Cell & Molecular Biology and the Plant Pathology Dept. communities for support and generosity. Thanks for giving me the opportunity to live and study in Madison - I had a truly wonderful time here. I also want to acknowledge the guidance of Dr. David Davido and the lab back at KU for all their training and support, as well as the biology and chemistry faculty from my undergraduate institute, Washburn University.

Thanks to Keppler, Tej, Kessens, Indro and Shaojie for all the badminton fun! Finally, I want to thank Grace for all the awesome baked goods and great company, and of course my family, who made this all possible.

## Table of Contents

Abstract .....	i
Acknowledgments.....	iii
Table of Contents .....	v
Chapter 1: Introduction .....	1
1.1 <i>Heterodera glycines</i> – the soybean cyst nematode (SCN).....	1
1.2 The <i>Rhg1</i> (Resistance to <i>Heterodera glycines</i> 1) locus .....	3
1.3 $\alpha$ -SNAP and NSF - the SNARE recycling machinery .....	5
1.4 Copia family LTR retrotransposons .....	8
1.5 Research Overview .....	10
1.6 References .....	11
Chapter 2: Copy number variation of multiple genes at <i>Rhg1</i> mediates nematode resistance in soybean .....	18
2.1 Abstract .....	19
2.2 Main Text .....	20
2.3 Materials and Methods .....	28
2.4 Acknowledgements .....	42
2.5 Figures.....	43
2.6 Supplemental Figures.....	49
2.7 Supplemental Tables .....	56
2.8 References .....	61
Chapter 3: Distinct copy number, coding sequence and locus methylation patterns underlie <i>Rhg1</i> -mediated soybean resistance to soybean cyst nematode.....	66
3.1 Abstract .....	67
3.2 Introduction .....	68
3.3 Results .....	73
3.4 Discussion .....	86
3.5 Materials and Methods .....	92
3.6 Acknowledgements .....	102
3.7 Figures.....	103
3.8 Tables .....	116

3.9 Supplemental Figures .....	118
3.10 Supplemental Tables .....	125
3.11 References .....	135
Chapter 4: Disease resistance through impairment of $\alpha$ -SNAP-NSF interaction and vesicular trafficking by soybean <i>Rhg1</i> .....	143
4.1 Abstract .....	144
4.2 Significance Statement .....	145
4.3 Introduction .....	146
4.4 Results .....	150
4.5 Discussion .....	160
4.6 Materials and Methods .....	165
4.7 Acknowledgements .....	172
4.8 Figures .....	173
4.9 Supporting Information .....	180
4.10 Supporting Tables .....	195
4.11 References .....	196
Chapter 5: An atypical NSF (N-ethylmaleimide Sensitive Factor) enables the viability of nematode-resistant <i>Rhg1</i> soybean .....	201
5.1 Abstract .....	202
5.2 Introduction .....	204
5.3 Results .....	211
5.4 Discussion .....	225
5.5 Materials and Methods .....	230
5.6 Acknowledgements .....	236
5.7 Figures .....	237
5.8 Tables .....	246
5.9 Supporting Information .....	247
5.10 Supplemental Tables .....	259
5.11 References .....	261

Chapter 6: An intact retrotransposon is integrated within the $\alpha$ -SNAP-encoding gene of low-copy haplotypes of the <i>Rhg1</i> locus .....	267
6.1 Abstract .....	268
6.2 Introduction .....	269
6.3 Results .....	273
6.4 Discussion .....	280
6.5 Materials and Methods .....	283
6.6 Acknowledgements .....	287
6.7 Figures .....	288
6.8 Supplemental Figures .....	294
6.9 References .....	298
Chapter 7: Future directions.....	302
7.1 Future Directions.....	302
7.2 Model .....	309
7.3 References .....	311

## Chapter 1: Introduction

### 1.1 *Heterodera glycines* – The soybean cyst nematode (SCN)

Plant parasitic nematodes damage many valuable crops including rice, beans, potato, tomato and wheat (Jones et al., 2013). Worldwide, annual yield losses attributed to plant parasitic nematodes value near 80 billion USD (Jones et al., 2013) Of these, the obligate sedentary parasites known as cyst nematodes are among the most harmful (Jones et al., 2013; Mitchum, 2016). Cyst nematodes parasite host roots via manipulating individual host cells to form an elaborate feeding structure, termed a syncytium (Kyndt et al., 2013). Cyst nematodes secrete molecules called effectors, which modulate host physiology and promote formation of the syncytial feeding structure (Hewezi and Baum, 2013; Mitchum et al., 2013). These effectors are essential for parasitism and are secreted from a protrusible mouth-spear called a stylet, which also enables feeding from the nematode-induced syncytium (Niblack et al., 2006).

The soybean cyst nematode (*Heterodera glycines*; SCN), is a highly adapted soybean root parasite and the most significant U.S. soybean pathogen, typically causing yield losses exceeding 1 billion USD each year (Niblack et al., 2006; Mitchum, 2016). Furthermore, SCN is present in all major US soybean growing regions as well as worldwide (Niblack et al., 2006; Mitchum, 2016). The SCN lifecycle consists of four juvenile stages (J1-J4) and an adult stage (sedentary female cyst or motile male) (Niblack et al., 2006). Upon hatching from soil-borne eggs, migratory J2 SCN seek out and enter host roots using stylet thrusts and the cell wall degrading effector secretions (Niblack et al., 2006). Once entering the root, J2 move to and select a single cell near the vasculature that is amenable to syncytium formation, and then pump distinct effector subsets into the cytoplasm to initiate host cell reprogramming (Kyndt et al., 2013). After syncytium formation begins, J2 become immobile and solely dependent on the

syncytium for nutrition (Niblack et al., 2006). Successful completion of the lifecycle depends on inducing and maintaining a functional syncytium for ~2-3 weeks for development into an adult female cyst or male (Niblack et al., 2006). Egg filled cysts die soon after fertilization and form a protective shell around the eggs, facilitating both dispersal and survival (Niblack et al., 2006).

Cysts in contaminated soil spread easily by wind, water or machinery. Furthermore, SCN hatching is elicited by environmental cues like host root signals, and unhatched eggs can remain viable for over a decade (Niblack et al., 2006). This makes cyst nematodes particularly difficult to eradicate once present in a field. Nematicidal treatments like fumigation may be used, but these strategies can be expensive and damaging to soil communities. Therefore, host resistance and crop rotation are the standard control strategies for SCN (Niblack et al., 2006; Mitchum, 2016).

The SCN-induced syncytium is a unique and highly complex organ (Kyndt et al., 2013). Biologically, syncytium formation is a fascinating cellular process entailing dissolution of cell walls, cell-cell fusion, endoreduplication, and drastic metabolic changes, among many events (de Almeida Engler and Gheysen, 2013; Hewezi and Baum, 2013; Kyndt et al., 2013). Likewise, the study of how nematode effectors accomplish these cellular manipulations is an intriguing research field. Many diverse and unique nematode effectors have been discovered, including ones that mimic plant signaling factors, alter host metabolism and interact with host transcription factors. (Bekal et al., 2003; Wang et al., 2010; Wang et al., 2011; Hewezi and Baum, 2013; Hewezi et al., 2015). Moreover, reports indicate that SCN even produces and secretes plant hormones such as cytokinins (Siddique et al., 2015).

The SCN-host interaction orchestrated by effector secretions is a fascinating biological process. Aside from transgenic soybean root models, other cyst nematode model systems including *Arabidopsis* and *Heterodera schachtii* are used and have greatly expanded our knowledge of syncytium formation and effector biology (Grunewald et al., 2009; Hewezi et al., 2010). Unfortunately, SCN research is hampered by the fact that no SCN reference genome is publicly available and little therefore is known about SCN genetics. A recent report generated SCN populations reared exclusive on SCN-resistant soybeans, providing an excellent resource for future genomic studies and effector variation (Gardner et al., 2017). Unfortunately, no techniques or methodologies current exist for stable transgenic manipulation of any plant parasitic nematode species and many studies of SCN effectors are therefore limited to expression *in planta* or silencing SCN effectors via host-induced gene silencing (Sindhu et al., 2009; Li et al., 2010; Hamamouch et al., 2012; Hewezi et al., 2015).

## **1.2 The *Rhg1* (Resistance to *Heterodera glycines* 1) locus**

SCN resistance is quantitative and *Glycine* species have evolved multiple loci to defend against SCN (Concibido et al., 2004). The *Rhg1* (Resistance to *Heterodera glycines* 1) locus, which was identified more than 50 years ago, is the major SCN resistance QTL and contributes to more than half of the total SCN resistance phenotype (Caldwell et al., 1960; Concibido et al., 2004; Mitchum, 2016). As such, all commercially available SCN-resistant soybeans utilize *Rhg1* as the primary SCN resistance locus (Cook et al., 2012; Cook et al., 2014).

*Rhg1* causes the developing SCN-induced syncytium to fail soon after initiation, thereby depriving the immobile nematode juvenile of sustenance (Mitchum, 2016). Initial studies misidentified the gene conferring *Rhg1* function for a LLR-like kinase (Lightfoot and Meksem,

2001). A subsequent study which silenced this kinase, however, showed no impacts on SCN resistance (Melito et al., 2010). Later studies investigating *Rhg1* function focused on transcriptomic changes occurring in syncytia of *Rhg1*-containing soybeans vs SCN-susceptible using laser capture microdissection (Klink et al., 2007; Ithal and Mitchum, 2011; Kandoth et al., 2011). These studies indicated that complex transcriptomic profiles, including oxidative stress response and unfolded protein responses, among others, change during the *Rhg1*-mediated response (Kandoth et al., 2011). The abundance of a transcript encoding a protein similar to AtBAG6, which is involved in cell death responses was also noted to be elevated in SCN-resistant syncytia (Kandoth et al., 2011).

In 2010, a genetic mapping study placed *Rhg1* within 11 putative candidate genes (Kim et al., 2010). The genes conferring *Rhg1*-phenotype were subsequently identified using gene silencing and SCN-complementation experiments by Cook *et al* in 2012 (Cook et al., 2012). This study found that the genes underlying *Rhg1* function do not resemble canonical plant disease resistance genes, but rather, *Rhg1* is a tandemly repeated block of four different genes: *Glyma.18G022400* (putative amino acid permease), *Glyma.18G022500* (putative  $\alpha$ -Soluble NSF Attachment Protein), *Glyma.18G022600* (PLAC8-domain) and *Glyma.18G022700* (putative wound inducible protein) (Cook et al., 2012). Additionally, two distinct classes or “alleles” of *Rhg1*, distinguished by SCN resistance phenotype had previously been reported by soybean breeders (Niblack et al., 2002; Brucker et al., 2005). In 2014, the natural variation at *Rhg1* which underlies these two phenotypic classes of *Rhg1* was examined (Cook et al., 2014). This study and that of Lee *et al* found that *Rhg1* loci group into two haplotypes based on *Rhg1* repeat copy number and on presence of distinctive  $\alpha$ -SNAP alleles (Cook et al., 2014; Lee et al., 2015). *Rhg1* high copy haplotypes contain at least 4 *Rhg1* blocks, one of which is a single WT  $\alpha$ -SNAP

encoding block, with the other repeats encoding “high copy type  $\alpha$ -SNAPs” (Cook et al., 2014). On the other hand, *Rhg1* low copy haplotypes possess three or fewer *Rhg1* blocks and all blocks encode a “low copy type  $\alpha$ -SNAP” (Cook et al., 2014). Notably, both types of *Rhg1* resistance type  $\alpha$ -SNAPs contain C-terminal amino acid polymorphisms at residues highly conserved in  $\alpha$ -SNAP encoded by other eukaryotes (Cook et al., 2014). Additional differences among *Rhg1* low copy vs. *Rhg1* high copy haplotypes in terms of mRNA transcript abundance and DNA methylation were also reported (Cook et al., 2014).

Most modern high yielding soybean varieties utilize a high copy type *Rhg1* derived from soybean accession PI 88788 (Niblack et al., 2008; Kim et al., 2010). However, near exclusive agricultural use of this *Rhg1* source is selecting for SCN populations with increased virulence, likely through different effector allelic combinations (Bekal et al., 2003; Gardner et al., 2017). Therefore, many studies have examined and reported additional QTL impacting SCN-resistance (Kopisch-Obuch and Diers, 2006; Liu et al., 2012; Vuong et al., 2015; Lakhssassi et al., 2017). Of these, a locus called *Rhg4*, is needed for full SCN resistance in *Rhg1* low copy, but not *Rhg1* high copy backgrounds (Brucker et al., 2005; Liu et al., 2012). Recently, the *Rhg4* locus was cloned and shown to encode a polymorphic serine hydroxy methyl transferase (SHMT) with altered kinetic properties (Liu et al., 2012). Additionally, a locus on chromosome 11 was reported as a minor SCN QTL (Lakhssassi et al., 2017).

### **1.3 $\alpha$ -SNAP and NSF - the SNARE recycling machinery**

Eukaryotic cells ship cargoes between various membrane compartments, as well as the cell exterior, using transport vesicles. For vesicle fusion to occur, cognate SNARE (Soluble N-ethylmaleimide Attachment protein REceptors) proteins on the vesicle and target membrane

surfaces must pair and “zipper” into SNARE bundles (SNARE complexes) that draw the membranes together (Jahn and Scheller, 2006; Wickner and Schekman, 2008). SNARE complexes are highly stable and complex formation does not require external energy inputs, however, to participate in new rounds of fusion, SNARE complexes must be forcefully disassembled into individual acceptor SNAREs (Jahn and Scheller, 2006).

The housekeeping proteins  $\alpha$ -Soluble NSF Attachment Protein ( $\alpha$ -SNAP) and N-ethylmaleimide Sensitive Factor (NSF) are the SNARE recycling machinery (Jahn and Scheller, 2006; Zhao et al., 2015). Together,  $\alpha$ -SNAP and NSF maintain vesicle trafficking by disassembling all SNARE complexes formed after vesicle fusion events (Zhao et al., 2015). Multiple  $\alpha$ -SNAPs bound to the SNARE bundle recruit and stimulate ATP hydrolysis by a NSF hexamer, which powers SNARE complex recycling. This large complex of four  $\alpha$ -SNAPs, a NSF hexamer and a SNARE bundle is known as the 20S complex (Zhao et al., 2015; Zhao and Brunger, 2016).

Given their central role in vesicle trafficking,  $\alpha$ -SNAP and NSF were identified many decades ago as key factors for secretion (Beckers et al., 1989; Clary et al., 1990; Kaiser and Schekman, 1990; Griff et al., 1992) Likewise, the structure and biochemical function of  $\alpha$ -SNAP and NSF is well known - crystal and cryo-EM structures have revealed  $\alpha$ -SNAP, SNARE-bundle and NSF contacts and configurations within the 20S supercomplex at high resolution (Rice and Brunger, 1999; Yu et al., 1999; Zhao et al., 2015) Many excellent reviews detailing SNARE-mediated vesicle fusion and SNARE recycling machinery are available (Zhao et al., 2007; Wickner and Schekman, 2008; Baker and Hughson, 2016; Zhao and Brunger, 2016).

NSF is a homo-hexameric ATPase belonging to the AAA+ family (ATPases associated with various cellular Activities) and couples ATP-hydrolysis with force-generating

conformational changes (Whiteheart et al., 2001; Hanson and Whiteheart, 2005). NSF contains 3 functional domains: the N-domain, the D1 ATPase and the D2 ATPase (Hanson and Whiteheart, 2005). The NSF N domain binds and maintains electrostatic contacts with  $\alpha$ -SNAP, while D1 ATPase domain drives SNARE disassembly (Zhao et al., 2015). The D2 ATPase has low intrinsic ATPase activity and mediates NSF multimerization (Zhao et al., 2010).  $\alpha$ -SNAP is an adaptor which promiscuously binds to conserved features of SNARE bundles (Marz et al., 2003). NSF requires  $\alpha$ -SNAP not only to bind SNAREs, but also to stimulate ATP hydrolysis and channel the conformational changes which unwind and remodel SNARE complexes (Zhao et al., 2015). The  $\alpha$ -SNAP C-terminus is conserved amongst diverse eukaryotes and directly interfaces with the NSF N-domain through complementary electrostatic contacts (Barnard et al., 1996, 1997). A conserved  $\alpha$ -SNAP C-terminal penultimate leucine is also implicated in stimulating NSF D1 domain ATPase activity (Barnard et al., 1996, 1997). The interactions of NSF and  $\alpha$ -SNAP with each other as well as the SNARE bundle, are coordinated through patches of electrostatic contacts (Zhao et al., 2015).

Only a handful of studies have examined  $\alpha$ -SNAP and NSF function in plants (Bassham and Raikhel, 1999; Rancour et al., 2002). However, mammalian NSF and  $\alpha$ -SNAP have been demonstrated to interact with plant SNAREs (Rancour et al., 2002). Further, functional conservation of both  $\alpha$ -SNAP and NSF has been demonstrated across kingdoms. Mammalian NSF will interact with plant  $\alpha$ -SNAPs and yeast homologs will complement mammalian systems (Clary et al., 1990; Bayless et al., 2016). This is likely due to conservation of electrostatic contacts occurring between  $\alpha$ -SNAP, NSF and the SNARE bundle (Marz et al., 2003; Zhao et al., 2015). Importantly, these studies indicate that the role of  $\alpha$ -SNAP and NSF as dedicated SNARE chaperones is conserved across eukaryotes.

Most animals encode single genes for  $\alpha$ -SNAP and NSF. Soybean, on the other hand, encodes 9 putative members of the SNARE-recycling machinery: 5  $\alpha$ -SNAPs, 2  $\gamma$ -SNAPs (Gamma) and 2 NSF genes (Schmutz et al., 2010). The precise role of gamma-SNAPs is unclear, but one study suggests gamma-SNAPs could function in SNARE disassembly at specific endomembrane networks (Inoue et al., 2015).

That unusual  $\alpha$ -SNAPs are involved in *Rhg1* mediated SCN resistance was therefore unexpected (Cook et al., 2012; Cook et al., 2014). Notably, *Rhg1*-encoded resistance-type  $\alpha$ -SNAPs are polymorphic at conserved C-terminal residues. It is these same C-terminal residues which function both to recruit and stimulate NSF (Barnard et al., 1996; Zhao et al., 2015). Apart from chaperoning SNARE disassembly, other roles for both  $\alpha$ -SNAP and NSF, are reported, such as binding of calcium channels and involvement in apoptotic cell death (Hanley et al., 2002; Zhao et al., 2007; Naydenov et al., 2012; Miao et al., 2013). Intriguingly, Sec17 (yeast  $\alpha$ -SNAP), was recently demonstrated to enhance the fusion of vesicles containing trans-paired SNAREs and a “proofreading” function in reduce non-cognate SNARE interactions was also proposed (Zick et al., 2015; Baker and Hughson, 2016).

#### **1.4 Copia family LTR retrotransposons**

While examining *Rhg1* structure, we unexpectedly discovered an intact 4.77 kb Copia retrotransposon within the  $\alpha$ -SNAP ORF. Curiously, this element was anti-sense to  $\alpha$ -SNAP and within intron 1.

Ty-1 Copia family retrotransposons earn their namesake for their often high abundance in plant genomes (Galindo-Gonzalez et al., 2017). Often regarded as genomic parasites which can disrupt important host genes, transposons, and retrotransposons are increasingly recognized as

drivers for genome evolution that impact chromatin structure, genome organization and impact nearby gene expression (McCue and Slotkin, 2012; Zhao and Ma, 2013; Mita and Boeke, 2016).

Two different LTR retrotransposon superfamilies - Gypsy and Copia - are known and characterized by gene order (Havecker et al., 2004). Autonomous retrotransposons contain a 5'LTR (Long terminal repeat), a polyprotein ORF, a 3' LTR and can initiate replication. The retrotransposon polyprotein contains several key domains including Gag, reverse transcriptase, integrase, and protease (Schulman, 2013). These retrotransposon proteins facilitate replication via copying a mRNA intermediate into DNA followed by integration at a new site within the host chromosome (Schulman, 2013). Transposons frequently integrate within euchromatic regions, adjacent or within host genes (Kidwell and Lisch, 2001; Lisch and Bennetzen, 2011).

Host cells therefore defend against unchecked transposable element expansion. These include DNA methylation of repetitive elements, and RNAi mechanisms (Lisch and Slotkin, 2011; Ito, 2012; Kim and Zilberman, 2014). Various stresses may compromise host surveillance and silencing of repetitive elements and allow element activation (Cavrak et al., 2014; Negi et al., 2016). However, stress-activation of transposons can be beneficial through effecting regulation of host genes (McCue et al., 2012; Le et al., 2015; Negi et al., 2016).

Transposable elements affecting host phenotypes was recognized long ago in landmark discoveries by McClintock (Mc, 1950). Since then, a plethora of reports has shown that transposons can impact various host genes broadly through epigenetic influences, but also in case specific instances that confer host benefits. Such reports include but are not limited to the following: flowering control, trichome production, disease resistance gene production and fruit size (Liu et al., 2004; Xiao et al., 2008; McCue et al., 2012; McCue and Slotkin, 2012; Tsuchiya and Eulgem, 2013; Ding et al., 2015).

## 1.5 Research Overview

This project investigates the molecular function of the major soybean cyst nematode resistance locus, *Rhg1*. In Chapter 2, the genes conferring *Rhg1* function, and the basic genetic structure of the *Rhg1* locus was identified— a minor role was performed in this chapter. Chapter 3 explored the natural diversity of various *Rhg1* sources and defined two distinctive resistance-conferring *Rhg1* haplotypes by copy number variation, unique  $\alpha$ -SNAP alleles, mRNA expression and methylation differences. Chapter 4 molecularly characterized the polymorphic *Rhg1*-resistance-type  $\alpha$ -SNAPs as well as their potential functions during the *Rhg1* SCN-resistance response. Chapter 5 extended our knowledge of the WT  $\alpha$ -SNAP balance in *Rhg1* varieties and identified that an usual NSF allele (RAN07) is necessary for the survival of multi-copy *Rhg1* germplasms. Moreover, the RAN07 polymorphisms were shown to biochemically and functionally enhance interactions with the *Rhg1* resistance-type  $\alpha$ -SNAPs. This study suggested that a broad rewiring of the core SNARE recycling machinery underlies SCN-resistance in soybean. Chapter 6 redefined our understanding of low copy *Rhg1* structure and revealed that an intact 4.8 kb retrotransposon was inserted within the resistance-type  $\alpha$ -SNAP locus. The potential effects of this Copia insertion on low copy  $\alpha$ -SNAP expression were then explored.

## 1.6 References

- Baker RW, Hughson FM** (2016) Chaperoning SNARE assembly and disassembly. *Nat Rev Mol Cell Biol* **17**: 465-479
- Barnard RJ, Morgan A, Burgoyne RD** (1996) Domains of alpha-SNAP required for the stimulation of exocytosis and for N-ethylmaleimide-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
- Bassham DC, Raikhel NV** (1999) The pre-vacuolar t-SNARE AtPEP12p forms a 20S complex that dissociates in the presence of ATP. *Plant J* **19**: 599-603
- Bayless AM, Smith JM, Song J, McMinn PH, Teillet A, August BK, Bent AF** (2016) Disease resistance through impairment of alpha-SNAP-NSF interaction and vesicular trafficking by soybean Rhg1. *Proc Natl Acad Sci U S A* **113**: E7375-E7382
- Beckers CJ, Block MR, Glick BS, Rothman JE, Balch WE** (1989) Vesicular transport between the endoplasmic reticulum and the Golgi stack requires the NEM-sensitive fusion protein. *Nature* **339**: 397-398
- Bekal S, Niblack TL, Lambert KN** (2003) A chorismate mutase from the soybean cyst nematode *Heterodera glycines* shows polymorphisms that correlate with virulence. *Mol Plant Microbe Interact* **16**: 439-446
- 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. *Theor Appl Genet* **111**: 44-49
- Caldwell BE, Brim C, Ross J** (1960) Inheritance of resistance of soybeans to the cyst nematode, *Heterodera glycines*. *Agronomy Journal* **52**: 635-636
- Cavrak VV, Lettner N, Jamge S, Kosarewicz A, Bayer LM, Mittelsten Scheid O** (2014) How a retrotransposon exploits the plant's heat stress response for its activation. *PLoS Genet* **10**: e1004115
- Clary DO, Griff IC, Rothman JE** (1990) SNAPs, a family of NSF attachment proteins involved in intracellular membrane fusion in animals and yeast. *Cell* **61**: 709-721
- Concibido VC, Diers BW, Arelli PR** (2004) A Decade of QTL Mapping for Cyst Nematode Resistance in Soybean. *Crop Sci.* **44**: 1121-1131

- Cook DE, Bayless AM, Wang K, Guo X, Song Q, Jiang J, Bent AF** (2014) Distinct Copy Number, Coding Sequence, and Locus Methylation Patterns Underlie Rhg1-Mediated Soybean Resistance to Soybean Cyst Nematode. *Plant Physiol* **165**: 630-647
- 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
- de Almeida Engler J, Gheysen G** (2013) Nematode-induced endoreduplication in plant host cells: why and how? *Mol Plant Microbe Interact* **26**: 17-24
- Ding M, Ye W, Lin L, He S, Du X, Chen A, Cao Y, Qin Y, Yang F, Jiang Y, Zhang H, Wang X, Paterson AH, Rong J** (2015) The Hairless Stem Phenotype of Cotton (*Gossypium barbadense*) Is Linked to a Copia-Like Retrotransposon Insertion in a Homeodomain-Leucine Zipper Gene (HD1). *Genetics* **201**: 143-154
- Galindo-Gonzalez L, Mhiri C, Deyholos MK, Grandbastien MA** (2017) LTR-retrotransposons in plants: Engines of evolution. *Gene* **626**: 14-25
- Gardner M, Heinz R, Wang J, Mitchum MG** (2017) Genetics and Adaptation of Soybean Cyst Nematode to Broad Spectrum Soybean Resistance. *G3 (Bethesda)* **7**: 835-841
- Griff IC, Schekman R, Rothman JE, Kaiser CA** (1992) The yeast SEC17 gene product is functionally equivalent to mammalian alpha-SNAP protein. *J Biol Chem* **267**: 12106-12115
- Grunewald W, Cannoot B, Friml J, Gheysen G** (2009) Parasitic nematodes modulate PIN-mediated auxin transport to facilitate infection. *PLoS Pathog* **5**: e1000266
- Hamamouch N, Li C, Hewezi T, Baum TJ, Mitchum MG, Hussey RS, Vodkin LO, Davis EL** (2012) The interaction of the novel 30C02 cyst nematode effector protein with a plant beta-1,3-endoglucanase may suppress host defence to promote parasitism. *J Exp Bot* **63**: 3683-3695
- Hanley JG, Khatri L, Hanson PI, Ziff EB** (2002) NSF ATPase and alpha-/beta-SNAPs disassemble the AMPA receptor-PICK1 complex. *Neuron* **34**: 53-67
- Hanson PI, Whiteheart SW** (2005) AAA+ proteins: have engine, will work. *Nat Rev Mol Cell Biol* **6**: 519-529
- Havecker ER, Gao X, Voytas DF** (2004) The diversity of LTR retrotransposons. *Genome Biol* **5**: 225
- Hwezi T, Baum TJ** (2013) Manipulation of plant cells by cyst and root-knot nematode effectors. *Mol Plant Microbe Interact* **26**: 9-16

- Hewezi T, Howe PJ, Maier TR, Hussey RS, Mitchum MG, Davis EL, Baum TJ** (2010) Arabidopsis spermidine synthase is targeted by an effector protein of the cyst nematode *Heterodera schachtii*. *Plant Physiol* **152**: 968-984
- Hewezi T, Juvale PS, Piya S, Maier TR, Rambani A, Rice JH, Mitchum MG, Davis EL, Hussey RS, Baum TJ** (2015) The cyst nematode effector protein 10A07 targets and recruits host posttranslational machinery to mediate its nuclear trafficking and to promote parasitism in Arabidopsis. *Plant Cell* **27**: 891-907
- Inoue H, Matsuzaki Y, Tanaka A, Hosoi K, Ichimura K, Arasaki K, Wakana Y, Asano K, Tanaka M, Okuzaki D, Yamamoto A, Tani K, Tagaya M** (2015) gamma-SNAP stimulates disassembly of endosomal SNARE complexes and regulates endocytic trafficking pathways. *J Cell Sci* **128**: 2781-2794
- Ithal N, Mitchum MG** (2011) Laser capture microdissection of nematode feeding cells. *Methods Mol Biol* **712**: 227-240
- Ito H** (2012) Small RNAs and transposon silencing in plants. *Dev Growth Differ* **54**: 100-107
- Jahn R, Scheller RH** (2006) SNAREs--engines for membrane fusion. *Nat Rev Mol Cell Biol* **7**: 631-643
- Jones JT, Haegeman A, Danchin EG, Gaur HS, Helder J, Jones MG, Kikuchi T, Manzanilla-Lopez R, Palomares-Rius JE, Wesemael WM, Perry RN** (2013) Top 10 plant-parasitic nematodes in molecular plant pathology. *Mol Plant Pathol* **14**: 946-961
- Kaiser CA, Schekman R** (1990) Distinct sets of SEC genes govern transport vesicle formation and fusion early in the secretory pathway. *Cell* **61**: 723-733
- Kandoth PK, Itahl 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 Physiol* **155**: 1960-1975
- Kidwell MG, Lisch DR** (2001) Perspective: transposable elements, parasitic DNA, and genome evolution. *Evolution* **55**: 1-24
- Kim M, Hyten DL, Bent AF, Diers BW** (2010) Fine Mapping of the SCN Resistance Locus rhg1-b from PI 88788. *Plant Gen.* **3**: 81-89
- Kim MY, Zilberman D** (2014) DNA methylation as a system of plant genomic immunity. *Trends Plant Sci* **19**: 320-326
- 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
- Kopisch-Obuch FJ, Diers BW** (2006) Segregation at the SCN resistance locus *rhg1* in soybean is distorted by an association between the resistance allele and reduced field emergence. *Theor Appl Genet* **112**: 199-207
- Kyndt T, Vieira P, Gheysen G, de Almeida-Engler J** (2013) Nematode feeding sites: unique organs in plant roots. *Planta* **238**: 807-818
- Lakhssassi N, Liu S, Bekal S, Zhou Z, Colantonio V, Lambert K, Barakat A, Meksem K** (2017) Characterization of the Soluble NSF Attachment Protein gene family identifies two members involved in additive resistance to a plant pathogen. *Sci Rep* **7**: 45226
- Le TN, Miyazaki Y, Takuno S, Saze H** (2015) Epigenetic regulation of intragenic transposable elements impacts gene transcription in *Arabidopsis thaliana*. *Nucleic Acids Res* **43**: 3911-3921
- Lee TG, Kumar I, Diers BW, Hudson ME** (2015) Evolution and selection of *Rhg1*, a copy-number variant nematode-resistance locus. *Mol Ecol* **24**: 1774-1791
- Li J, Todd TC, Oakley TR, Lee J, Trick HN** (2010) Host-derived suppression of nematode reproductive and fitness genes decreases fecundity of *Heterodera glycines* Ichinohe. *Planta* **232**: 775-785
- Lightfoot D, Meksem K** (2001) Isolated polynucleotides and polypeptides relating to loci underlying resistance to soybean cyst nematode and soybean sudden death syndrome and methods employing same. *In*. Google Patents
- Lisch D, Bennetzen JL** (2011) Transposable element origins of epigenetic gene regulation. *Curr Opin Plant Biol* **14**: 156-161
- Lisch D, Slotkin RK** (2011) Strategies for silencing and escape: the ancient struggle between transposable elements and their hosts. *Int Rev Cell Mol Biol* **292**: 119-152
- Liu J, He Y, Amasino R, Chen X** (2004) siRNAs targeting an intronic transposon in the regulation of natural flowering behavior in *Arabidopsis*. *Genes Dev* **18**: 2873-2878
- Liu S, Kandoth PK, Warren SD, Yeckel G, Heinz R, Alden J, Yang C, Jamai A, El-Mellouki T, Juvalle 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
- Marz KE, Lauer JM, Hanson PI** (2003) Defining the SNARE complex binding surface of alpha-SNAP: implications for SNARE complex disassembly. *J Biol Chem* **278**: 27000-27008

- Mc CB** (1950) The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci U S A* **36**: 344-355
- McCue AD, Nuthikattu S, Reeder SH, Slotkin RK** (2012) Gene expression and stress response mediated by the epigenetic regulation of a transposable element small RNA. *PLoS Genet* **8**: e1002474
- McCue AD, Slotkin RK** (2012) Transposable element small RNAs as regulators of gene expression. *Trends Genet* **28**: 616-623
- 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 Biol* **10**: 104
- Miao Y, Miner C, Zhang L, Hanson PI, Dani A, Vig M** (2013) An essential and NSF independent role for alpha-SNAP in store-operated calcium entry. *Elife* **2**: e00802
- Mita P, Boeke JD** (2016) How retrotransposons shape genome regulation. *Curr Opin Genet Dev* **37**: 90-100
- Mitchum MG** (2016) Soybean Resistance to the Soybean Cyst Nematode *Heterodera glycines*: An Update. *Phytopathology* **106**: 1444-1450
- Mitchum MG, Hussey RS, Baum TJ, Wang X, Elling AA, Wubben M, Davis EL** (2013) Nematode effector proteins: an emerging paradigm of parasitism. *New Phytol*
- Naydenov NG, Harris G, Brown B, Schaefer KL, Das SK, Fisher PB, Ivanov AI** (2012) Loss of soluble N-ethylmaleimide-sensitive factor attachment protein alpha (alphaSNAP) induces epithelial cell apoptosis via down-regulation of Bcl-2 expression and disruption of the Golgi. *J Biol Chem* **287**: 5928-5941
- Negi P, Rai AN, Suprasanna P** (2016) Moving through the Stressed Genome: Emerging Regulatory Roles for Transposons in Plant Stress Response. *Front Plant Sci* **7**: 1448
- Niblack T, Colgrove A, Colgrove K, Bond J** (2008) Shift in virulence of soybean cyst nematode is associated with use of resistance from PI 88788. *Plant Health Prog* **10**
- Niblack TL, Arelli PR, Noel GR, Opperman CH, Orf JH, Schmitt DP, Shannon JG, Tylka GL** (2002) A Revised Classification Scheme for Genetically Diverse Populations of *Heterodera glycines*. *J Nematol* **34**: 279-288
- Niblack TL, Lambert KN, Tylka GL** (2006) A model plant pathogen from the kingdom Animalia: *Heterodera glycines*, the soybean cyst nematode. *Annu Rev Phytopathol* **44**: 283-303

- Rancour DM, Dickey CE, Park S, Bednarek SY** (2002) Characterization of AtCDC48. Evidence for multiple membrane fusion mechanisms at the plane of cell division in plants. *Plant Physiol* **130**: 1241-1253
- 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
- Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, 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. *Nature* **463**: 178-183
- Schulman AH** (2013) Retrotransposon replication in plants. *Curr Opin Virol* **3**: 604-614
- Siddique S, Radakovic ZS, De La Torre CM, Chronis D, Novak O, Ramireddy E, Holbein J, Matera C, Hutten M, Gutbrod P, Anjam MS, Rozanska E, Habash S, Elashry A, Sobczak M, Kakimoto T, Strnad M, Schmulling T, Mitchum MG, Grundler FM** (2015) A parasitic nematode releases cytokinin that controls cell division and orchestrates feeding site formation in host plants. *Proc Natl Acad Sci U S A* **112**: 12669-12674
- 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. *J Exp Bot* **60**: 315-324
- Tsuchiya T, Eulgem T** (2013) An alternative polyadenylation mechanism coopted to the Arabidopsis RPP7 gene through intronic retrotransposon domestication. *Proc Natl Acad Sci U S A* **110**: E3535-3543
- Vuong TD, Sonah H, Meinhardt CG, Deshmukh R, Kadam S, Nelson RL, Shannon JG, Nguyen HT** (2015) Genetic architecture of cyst nematode resistance revealed by genome-wide association study in soybean. *BMC Genomics* **16**: 593
- Wang J, Lee C, Replogle A, Joshi S, Korkin D, Hussey R, Baum TJ, Davis EL, Wang X, Mitchum MG** (2010) Dual roles for the variable domain in protein trafficking and host-specific recognition of Heterodera glycines CLE effector proteins. *New Phytol* **187**: 1003-1017
- Wang J, Replogle A, Hussey R, Baum T, Wang X, Davis EL, Mitchum MG** (2011) Identification of potential host plant mimics of CLAVATA3/ESR (CLE)-like peptides from the plant-parasitic nematode Heterodera schachtii. *Mol Plant Pathol* **12**: 177-186
- Whiteheart SW, Schraw T, Matveeva EA** (2001) N-ethylmaleimide sensitive factor (NSF) structure and function. *Int Rev Cytol* **207**: 71-112

- Wickner W, Schekman R** (2008) Membrane fusion. *Nat Struct Mol Biol* **15**: 658-664
- Xiao H, Jiang N, Schaffner E, Stockinger EJ, van der Knaap E** (2008) A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science* **319**: 1527-1530
- Yu RC, Jahn R, Brunger AT** (1999) NSF N-terminal domain crystal structure: models of NSF function. *Mol Cell* **4**: 97-107
- Zhao C, Matveeva EA, Ren Q, Whiteheart SW** (2010) Dissecting the N-ethylmaleimide-sensitive factor: required elements of the N and D1 domains. *J Biol Chem* **285**: 761-772
- Zhao C, Slevin JT, Whiteheart SW** (2007) Cellular functions of NSF: not just SNAPs and SNAREs. *FEBS Lett* **581**: 2140-2149
- Zhao M, Brunger AT** (2016) Recent Advances in Deciphering the Structure and Molecular Mechanism of the AAA+ ATPase N-Ethylmaleimide-Sensitive Factor (NSF). *J Mol Biol* **428**: 1912-1926
- Zhao M, Ma J** (2013) Co-evolution of plant LTR-retrotransposons and their host genomes. *Protein Cell* **4**: 493-501
- Zhao M, Wu S, Zhou Q, Vivona S, Cipriano DJ, Cheng Y, Brunger AT** (2015) Mechanistic insights into the recycling machine of the SNARE complex. *Nature* **518**: 61-67
- Zick M, Orr A, Schwartz ML, Merz AJ, Wickner WT** (2015) Sec17 can trigger fusion of trans-SNARE paired membranes without Sec18. *Proc Natl Acad Sci U S A* **112**: E2290-2297

## **Chapter 2: Copy number variation of multiple genes at *Rhg1* mediates nematode resistance in soybean**

This chapter was previously published as:

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

Contributions: David Cook, Tong Geon Leon and Xiao-li Guo contributed the majority of the described work (co-first authors). I was a minor contributor and generated two RNAi silencing constructs for 2570 and 2620 used in Fig. 1B, and the 3 gene over-expression construct used in Fig. S4B as well as conceptual contributions in the final months of the project.

## 2.1 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 multiple genes at *rhg1-b*, encoding an amino acid transporter, an  $\alpha$ -SNAP protein and a protein with a WI12 (wound-inducible) domain, each contribute to resistance. These genes are present in one copy per haploid genome in susceptible varieties, but in ten tandem copies of a 31 kilobase segment 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.

## 2.2 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, 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 spread to all major U.S. soybean-producing states, and once it is present in a field there are no practical means of eradication (Niblack et al., 2006).

SCN molts through multiple juvenile life stages, including obligate endoparasitic stages on plant roots, to complete its life cycle (Niblack et al., 2006). The infective J2 juvenile stage invades roots of both susceptible and resistant soybean hosts. During feeding site establishment adjacent to the root vascular cylinder, root cells are reprogrammed through a highly evolved interaction facilitated by secreted nematode effectors (Davis et al., 2008; Gheysen and Mitchum, 2011). Many potential feeding sites are not formed, senesce, and/or are not maintained in *Rhgl*-mediated resistance (Niblack et al., 2006).

The *Rhgl* (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). Roughly 90% of the commercially cultivated soybean varieties marketed as SCN-resistant in the central U.S. use the *rhgl-b* allele (haplotype), derived from *G. max* PI 88788, as the primary SCN resistance locus. The molecular basis of this SCN-resistance has remained unclear.

Genetic mapping has placed *rhgl-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). An amino acid polymorphism in the *Glyma18g02590*-encoded  $\alpha$ -SNAP protein in this interval was recently suggested to contribute to SCN resistance (Matsye et al., 2012), although the authors indicated that this polymorphism does not account for *rhg1-b*-mediated resistance. A former *Rhg1* candidate gene, encoding an LRR-kinase protein, maps outside of the *rhg1-b* genetic interval and did not have a detectable impact on SCN resistance in transgenic root experiments (Kim et al., 2010; Melito et al., 2010). None of the gene products within the *rhg1-b* genetic interval resembles a canonical plant immune receptor (PRR or R gene product) (Dodds and Rathjen, 2010).

Phenotypic variation is often attributed to single nucleotide polymorphisms, but growing evidence from metazoa and vascular plants suggests that genome structural variation (SV) is a frequent and powerful driver of phenotypic diversity (Sebat et al., 2004; Springer et al., 2009). Copy number variation (CNV), a type of SV, has been shown to impact gene expression (Stranger et al., 2007) and contributes to a number of adaptive traits in humans, plants and insects (Schmidt et al., 2010; Stankiewicz and Lupski, 2010; Pearce et al., 2011; Wingen et al., 2012). 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 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 of other genes in and around the locus did not impact SCN resistance (e.g., Fig 1B, genes *Glyma18g02570* and *2620*). The three *Rhg1* genes discovered to contribute to SCN resistance encode a predicted amino acid transporter (*Glyma18g02580*), an  $\alpha$ -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 known 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 three genes that contribute to SCN resistance, 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 a 3' fragment of *Glyma18g02570* is immediately adjacent to the intergenic sequence downstream of (centromeric to) *Glyma18g02610* (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.

Sequencing and PCR amplification confirmed the presence of the *Glyma18g02610-2570* junction in DNA from multiple SCN-resistant soybean accessions, including accessions that carry the commercially important (Niblack et al., 2006) PI 88788, Peking and PI 437654 haplotypes of the *Rhg1* locus (Fig. 2C and fig. S2). The junction was not detected in four tested SCN-susceptible varieties including Williams 82 (fig. S2). This constitutes a direct test for

economically desirable alleles of the *Rhg1* locus. The shared identity of the junction sites from 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 impact SCN resistance exhibit significantly more transcript abundance in roots of SCN-resistant varieties relative to susceptible lines (Fig. 2D and fig. S2). In contrast, the transcript abundance for genes immediately flanking the SCN-impacting genes did not differ significantly between SCN-resistant and susceptible varieties (Fig. 2; *Glyma18g02600* expression in roots is at or below the limits of detection of qPCR, cDNA cloning and RNAseq methods (Severin et al., 2010; M.M., 2012)). Full-length transcripts were confirmed for *Glyma18g02580*, -2590 and -2610, and no hybrid repeat-junction transcripts were detected for *Glyma18g02570* (18; 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 utilized to directly determine the number of copies and arrangement of the 31 kb repeat segment in different haplotypes of the *Rhg1* locus. The hybridization pattern and DNA fiber length estimates generated using these probes (Fig. 3 and table S1) are consistent with the presence of 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 predicted 31 kb repeat segment, in the same configuration throughout the multiple nuclei sampled, in a pattern indicating ten direct repeats abutting in a head-to-tail arrangement (Fig. 3 and table S1). No additional copies (e.g., at other loci) were evident. In samples from soybean line Peking, three copies per DNA fiber were present in apparent direct repeat orientation (Fig. 3). Although fiber-FISH cannot resolve small

sequence differences, the single size of all junction-amplification PCR products and the consistency of all junction sequences assembled from fosmid or genomic DNA sequencing (Fig. S2) further suggest the presence of adjacent direct repeat copies.

The source of the first duplication event to arise at *Rhg1* is not known, but was possibly the result of nearby Ty1/copia-like retrotransposon RTvr1 or RTvr2 activity (Bennetzen, 2000; M.M., 2012). Later copy number expansion may have occurred by rare unequal exchange events between homologous repeats during meiotic recombination.

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  $\alpha$ -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 the set of genes within the 31 kb repeat segment was tested as a possible source of SCN resistance. A single recombinant DNA construct was made in which each of the genes *Glyma18g02580*, *-2590*, *-2600* and *-2610* was fused to a strong promoter. In two separate experiments that together tested >25 independent transgenic

events for each DNA construct, resistance to SCN was significantly increased in SCN-susceptible Williams 82 by simultaneous overexpression of this set of genes (Fig. 4). Increased SCN resistance was conferred despite the fact that three of the genes being overexpressed encode predicted amino acid products identical to those of SCN-susceptible Williams 82, and the polymorphic Fayette *rhg1-b Glyma18g02590* gene that was used was not sufficient to cause a detectable change in SCN resistance when overexpressed on its own (Fig. 4). Of note, there was no significant elevation of *PR-1* in these transgenic roots, which could have indicated non-specific elevation of defenses (fig. S4).

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 knowledge suggests future approaches to enhance the efficacy of *Rhg1*-mediated quantitative resistance to the highly important SCN disease of soybean, for example through isolation of soybean lines that carry more copies of the 31 kb *Rhg1* repeat, or through transgenic overexpression of the relevant genes. The latter approach may be applicable in other species as well, 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 using Phyre2 (M.M., 2012) indicated, 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. Hence Glyma18g02610 may participate in the production of phenazine-like compounds that are toxic to nematodes. Secretion of the Glyma18g02610 protein or other plant products that contribute to

disease resistance may be impacted by the Glyma18g02590  $\alpha$ -SNAP protein (Frei dit Frey and Robatzek, 2009). Because it is one of at least five  $\alpha$ -SNAP homologs encoded in the soybean reference genome, Glyma18g02590 may have undergone subfunctionalization or neofunctionalization (Conrad and Antonarakis, 2007). Fully sequenced plant genomes carry from two dozen to over five dozen annotated amino acid transporters of many subtypes ([www.phytzome.net](http://www.phytzome.net)), which can be involved in amino acid import and/or export between cells or between subcellular organelles (Okumoto and Pilot, 2011). The Glyma18g02580 protein and its most closely related transporters of soybean and other species are not functionally well-characterized, so the concept that Glyma18g02580 alters nematode success by altering the levels of specific amino acids or amino acid derivatives at the feeding site is only one of many viable hypotheses for future study regarding the SCN-deterring function of Glyma18g02580.

CNV of a block of dissimilar genes, rather than CNV for a single gene family, confers *Rhg1*-mediated SCN resistance. Recent analyses of genome-architecture in sorghum, rice, and soybean have reported high levels of CNV, and a tendency for overlap of regions of CNV 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 CNV conferring a valuable disease resistance trait. In humans and insects, adaptive traits have been associated with CNV for specific single genes (Perry et al., 2007; Schmidt et al., 2010). Single-copy clusters of functionally related but non-homologous genes are highly unusual in multicellular eukaryotes, but these have been reported in association with plant secondary metabolism (Field et al., 2011; Winzer et al., 2012). We provide a unique example of CNV involving more than two repeats, with the repeat encoding multiple gene products that are necessary for adaptation to the same important environmental constraint. Given the highly repetitive nature and plasticity of plant

genomes and the relatively underexplored association between CNV and phenotypes, it seems likely that a number of other complex traits are controlled by this type of CNV.

## 2.3 Materials and Methods

### *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 12N 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, 1X Gamborg's B5 vitamins, 7% Noble agar, pH 5.8) in 100 x 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<sub>600</sub> 0.6 - 0.7 in sterile ddH<sub>2</sub>O). The cotyledons were then placed abaxial-side down on a co-culture medium (CCM) (0.31g/L Gamborg's B5 salts, 3% sucrose, 1X Gamborg's B5 vitamins (BioWorld, Dublin OH), 0.4g/L L-cysteine, 0.154g/L dithiothreitol, 0.245g/L sodium thiosulfate, 40mg/L acetosyringone, 5% Noble agar, pH 5.4) in 100 x 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 4.3g/L MS salts (Sigma Co., St. Louis, MO), 2% sucrose, 1X

Gamborg's B5 vitamins (BioWorld, Dublin, OH), 7% Noble agar, 0.15g/L cefotaxime, 0.15g/L carbenicillin, pH 5.6 in 100 x 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.

### **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 710µm - 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  $\mu\text{m}$  sieve. Collected cysts were stored at 4°C in sealable plastic bags containing twice-sterilized flint sand.

### **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  $\mu\text{m}$  - 75  $\mu\text{m}$  - 25  $\mu\text{m}$  sieves (USA Standard Testing Sieve). Eggs were collected from the 25  $\mu\text{m}$  sieve and rinsed. Eggs were placed in a hatch chamber (Wong et al., 1993) with 3 mM  $\text{ZnCl}_2$  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 (Bybd 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 ( $[(J3 + \text{adult males} + \text{adult females}) / (J2 + J3 + \text{adult males} + \text{adult females})]$ ). 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).

### **Primer Table**

Primer sequences used to perform this research are listed in Supplemental Table 3 and referred to by number in this document.

### **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 300-600bp 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 AvrII/AscI (forward primer) and BamHI/SwaI (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 SwaI and AscI 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 BamHI and AvrII. 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, 6X 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 XbaI/KpnI and ligated to yield

*GmUbi<sub>prom</sub>:2610-HA:NOS<sub>term</sub>* (OE:2610-HA). The same procedure was used for *Glyma18g02590* (Table S3 57, 59), except the amplicon contained XbaI/SalI sites and was TA cloned into pCR8. *2590-HA:NOS<sub>term</sub>* and pSM103 were digested with XbaI/SalI and ligated to yield *GmUbi<sub>prom</sub>:2590-HA:NOS<sub>term</sub>* (OE:2590-HA). The full OE:2590-HA was also digested (XbaI/SalI) and ligated into pSM103 containing OE:2610-HA to yield OE:2610-OE:2590. To generate the four gene overexpression construct, the restriction sites PacI, PspOMI, and AscI were added to pSM101 between sites PstI/HindII 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 PstI/KpnI 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<sub>prom</sub>:2600-myc:NOS<sub>term</sub>* (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 HinIII/AscI 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<sub>prom</sub>:2580-myc:NOS<sub>term</sub>* (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<sub>FayP</sub>:2590<sub>Fay</sub>*) 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 SalI digestion and cloned into pSM101 using the SalI

restriction site. This sequence contained approximately 1kb of 5' regulatory DNA sequence. An additional 600bp 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 HindIII/Sall. 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.

### **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 ug to 1.0 ug 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 -80C. 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 -80C. 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

spectrophotometer (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 BioAnalyzer (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  $\mu$ M and 0.3  $\mu$ M. 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 X-X. 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 analysis (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).

### **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 185 bp sequence with 75% identity to the 5' and 3' long terminal repeat (LTR) regions of Ty1/copia-like retrotransposons RTvr1 and RTvr2 is present within 400 bp of the rhg1-b duplication junction.

### **Statistical analysis**

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

### **Fosmid library construction**

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 photocycle 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%  $\beta$ -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 ul of TE (10 mM Tris pH 7.5, 1mM EDTA). The DNA was treated with RNase A by incubating in 20 ug/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 ug 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<sup>R</sup> *E. coli* strain.

### **Fosmid clone sequencing and assembly**

Five candidate fosmid clones were identified by PCR-based pool screening using primers based on the *rhg1-b* interval of the Williams 82 reference sequence. Once it was confirmed that end sequences matched the anticipated region of the reference soybean genome sequence, they were sequenced using both the Roche 454/GS FLX+ system (Roche) and Illumina MiSeq (Illumina). 1-3 ug of fosmid clone DNA was used for making paired-end sequencing libraries for 454/GS FLX+. After library construction, pooled barcoded libraries were loaded onto one lane of the sequencing flow cell and sequenced. The average read length was 463 bp. The number of reads generated from 454/GS FLX+ is as follows: fosmid clone #1 in Fig. 2A: 10,865, #2: 6,271, #3:

6,648, #4: 6,520, and #5: 9,390. The reads were assembled using Phrap/Cross\_match (www.phrap.org) and CAP3 (Huang et al., 2006). For the MiSeq, 0.3-2 ug of DNA was used for making the sequencing library. Average DNA fragment size was 550 bp (range from 430 to 720 bp). 154 cycles from each end of the fragments were performed using a TruSeq SBS sequencing kit version 1 and analyzed with Casava1.8 (pipeline 1.8). Throughout the reads, the average quality scores for each base were over 30. The number of reads generated from MiSeq is as follows: fosmid clone #1 in Fig. 2A: 1,067,403, #2: 814,728, #3: 1,156,784, #4: 1,091,852, and #5: 946,028. ABySS (Simpson et al., 2009) was used to assemble the reads from MiSeq. The result was visualized using Geneious. Homopolymeric sequences and other problematic regions were manually sequenced using Sanger primer walking.

### **Whole-genome shotgun 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 ug 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  $28 \times$  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.

### **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/ $\mu$ m conversion rate (Cheng et al., 2002).

## 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. 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.

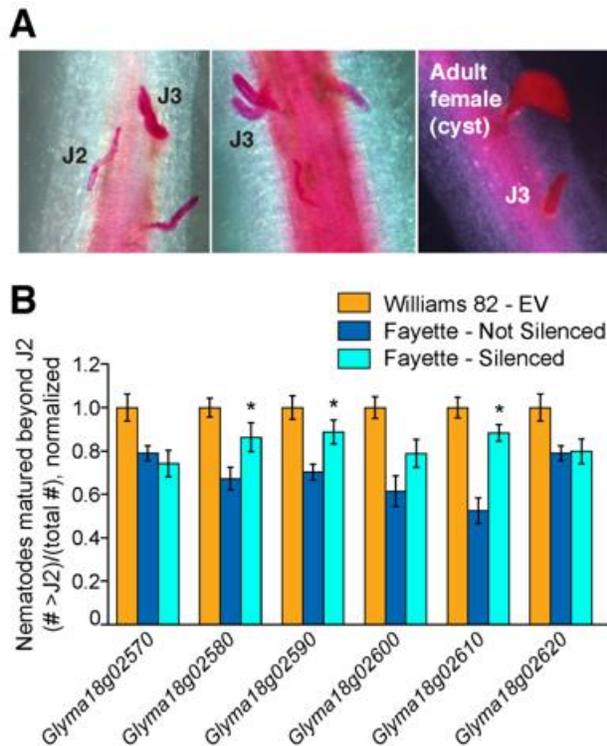
## Protein structure prediction and comparison

The protein structure for the predicted Glyma18g02610 gene product was modeled and proteins with the most homologous structures were identified using Phyre2 (Kelley and Sternberg, 2009), with default settings.

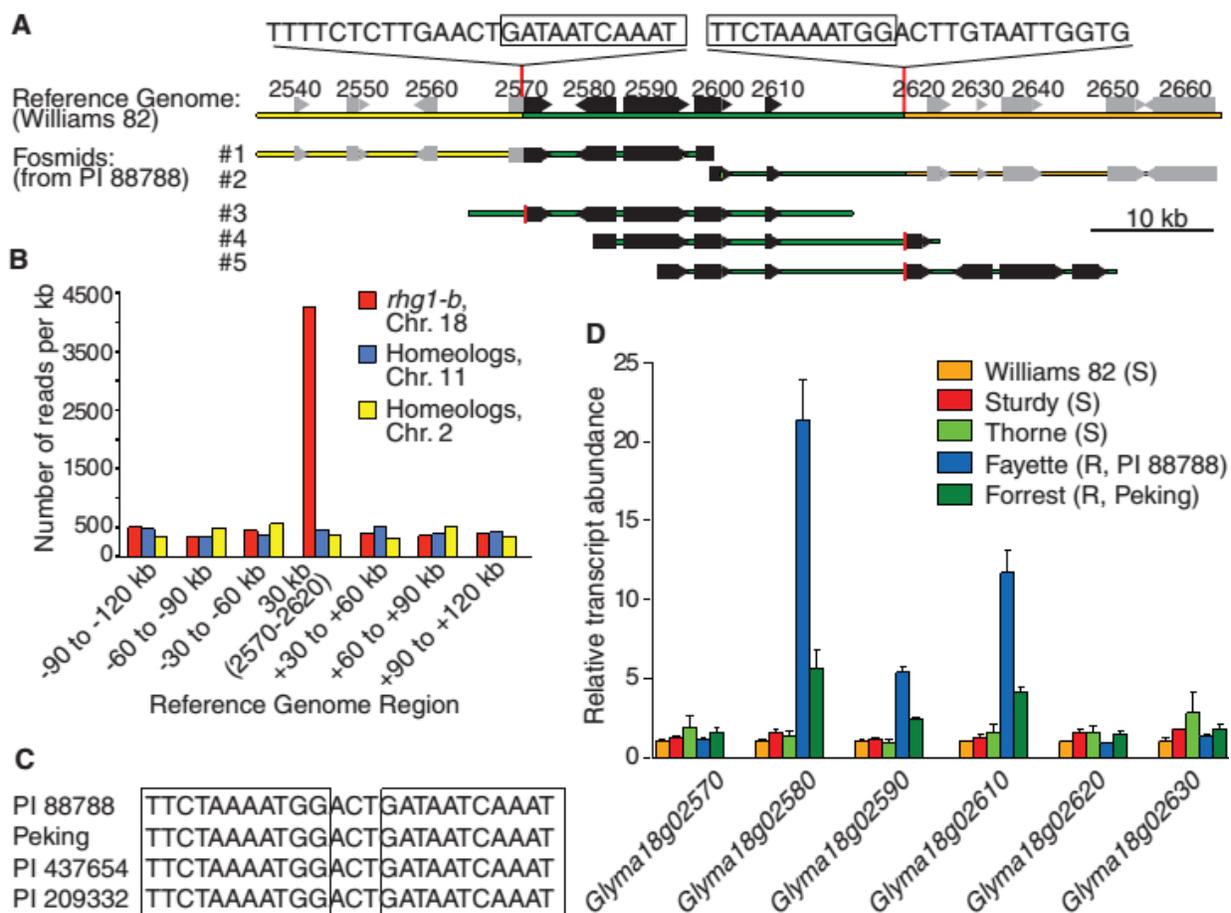
## **2.4 Acknowledgements**

This work was primarily supported by a United Soybean Board project award to A.F.B., a Illinois-Missouri Biotechnology Alliance award to M.H. and B.W.D, and National Science Foundation grant DBI-0922703 to J.J. Additional support came from a Wisconsin Experiment Station Hatch Award to A.F.B. and a 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.

## 2.5 Figures

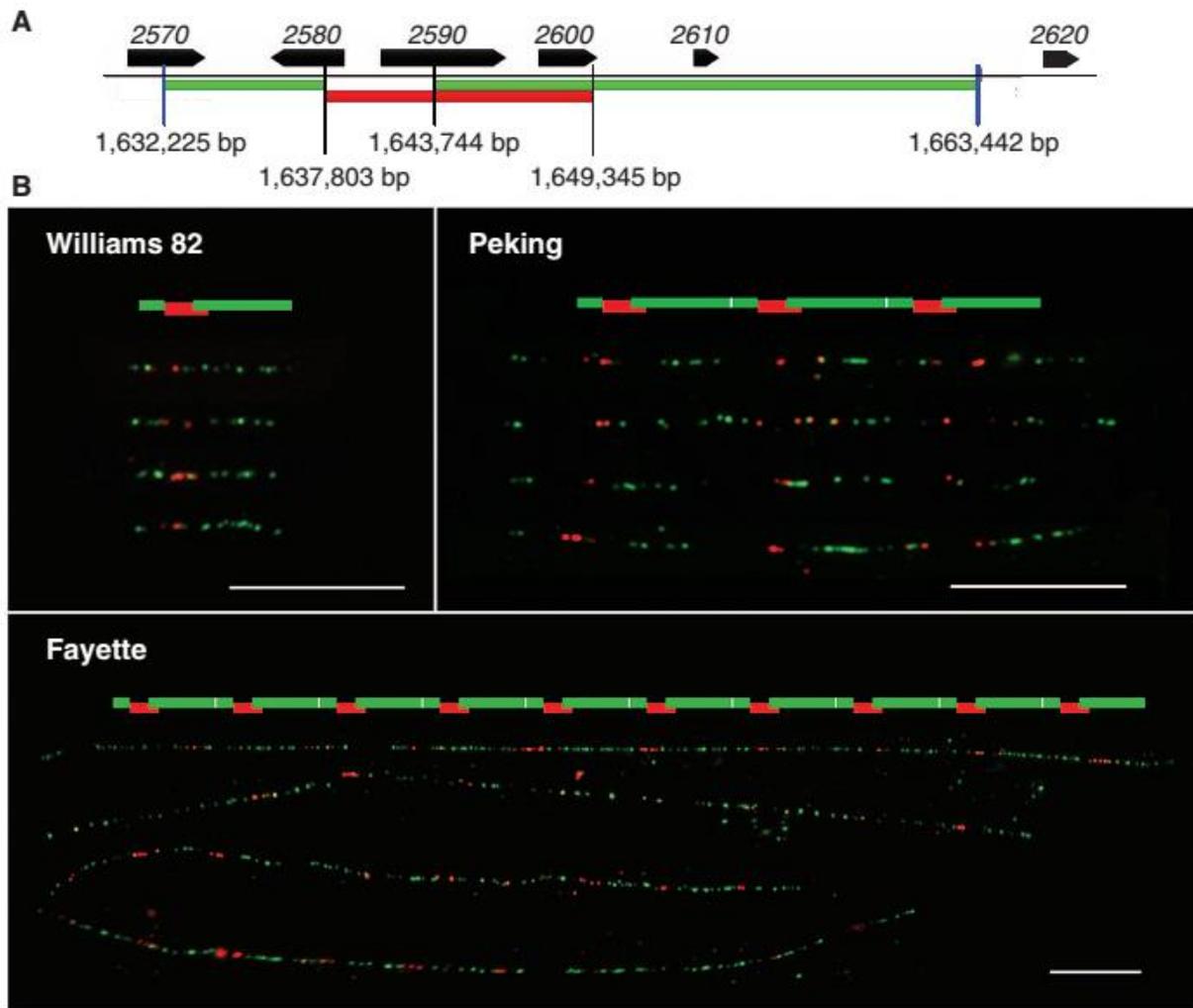


**Fig. 1.** Three genes at *rhg1-b* contribute to SCN resistance. **(A)** Representative SCN-infested roots; root vascular cylinder and nematodes stained with acid fuchsin. Fewer nematodes progress from J2 to J3, J4, adult male or egg-filled adult female (cyst) stages in SCN-resistant roots. **(B)** 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  $\pm$  std. error of mean. \*: Fayette (silenced) significantly different from Fayette (not silenced) based on ANOVA  $p < 0.05$ . EV: transformed with empty vector.



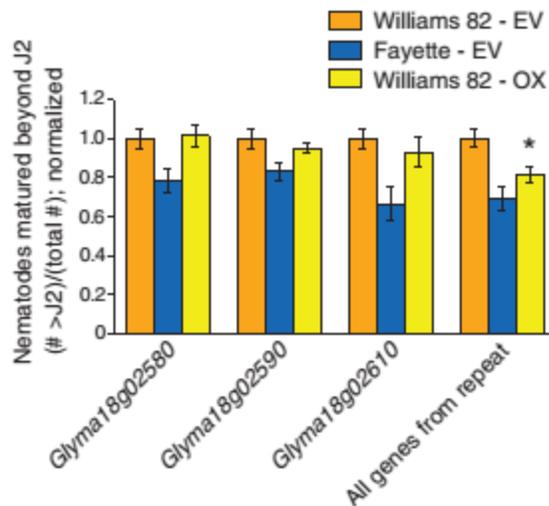
**Fig. 2.** A 31.2 kb repeat that elevates expression of the encoded genes is present in SCN-resistant haplotypes of the *Rhg1* locus. **(A)** Schematic of *Rhg1* locus of Williams 82 (top), and five fosmid inserts from *rhg1-b* haplotype. DNA sequences of soybean reference genome shown for the two designated locations. 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)** *Rhg1* repeat junction sequence from four different sources of SCN resistance (compare to reference genome sequences in (A)). **(C)** Number of whole-genome shotgun sequencing reads corresponding to reference genome region shown in green in (A) was ten-fold greater than for genome regions adjacent to *rhg1-b* on chromosome 18 or for *Rhg1*-homeologous loci on chromosomes 11 and 2.

(D) Transcript abundance of genes encoded in the 31 kb repeat region is much greater in roots from SCN-resistant soybean varieties relative to SCN-susceptible varieties. Mean  $\pm$  std. error of mean shown for qPCR; results for *Glyma18g02600* were at limit of detection.



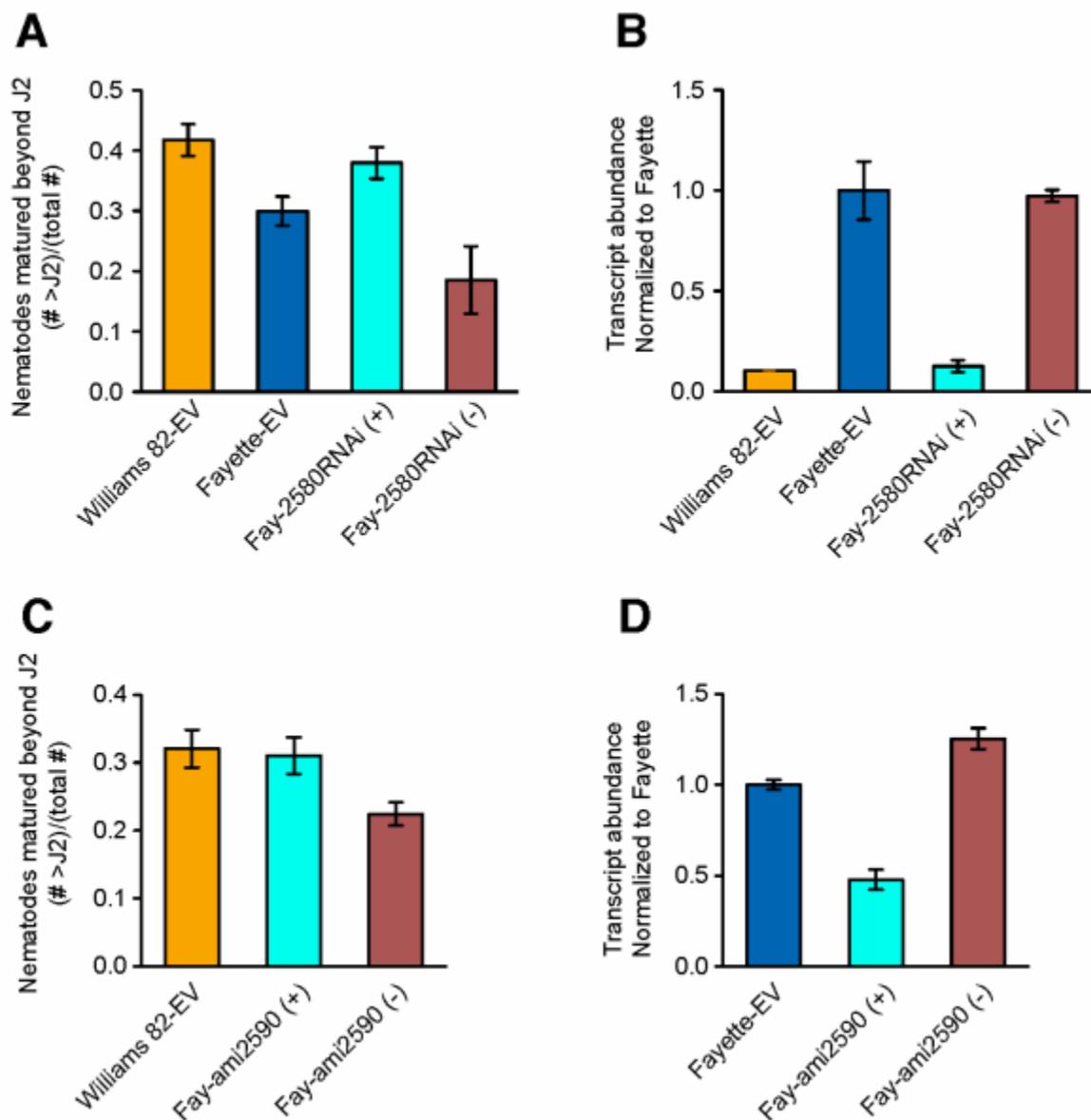
**Fig. 3.** Fiber-FISH detection of *Rhg1* copy number variation in widely used soybean lines. **(A)** Two adjacent probes were isolated from a single PI88788 (*rhg1-b*) genomic DNA fosmid clone whose insert spans a repeat junction, generating a 25.2 kb probe (green label) and an adjacent 9.7 kb probe (red label) as shown in Fig. 3A. DNA for green-labeled and red-labeled fiber-FISH probes are shown under the corresponding sequence regions of Williams 82. The 25.2 kb fragment from *rhg1-b* haplotype used for green probe was a single continuous DNA fragment that spans a repeat junction. **(B)** Composite of four Fiber-FISH images (four DNA fibers) per genotype, and probe diagram. Alternating pattern of red and green hybridization on single

genomic DNA fibers indicates ten and three direct repeat copies of the 31 kb block at *Rhg1* locus of SCN-resistant Fayette (*rhg1-b* derived from PI 88788) and Peking (PI 548402) respectively, and one copy per *Rhg1* haplotype in SCN-susceptible Williams 82. White bars = 10  $\mu\text{m}$ , which correspond to approximately 32 kb using a 3.21 kb/ $\mu\text{m}$  conversion rate (Cheng et al., 2002).



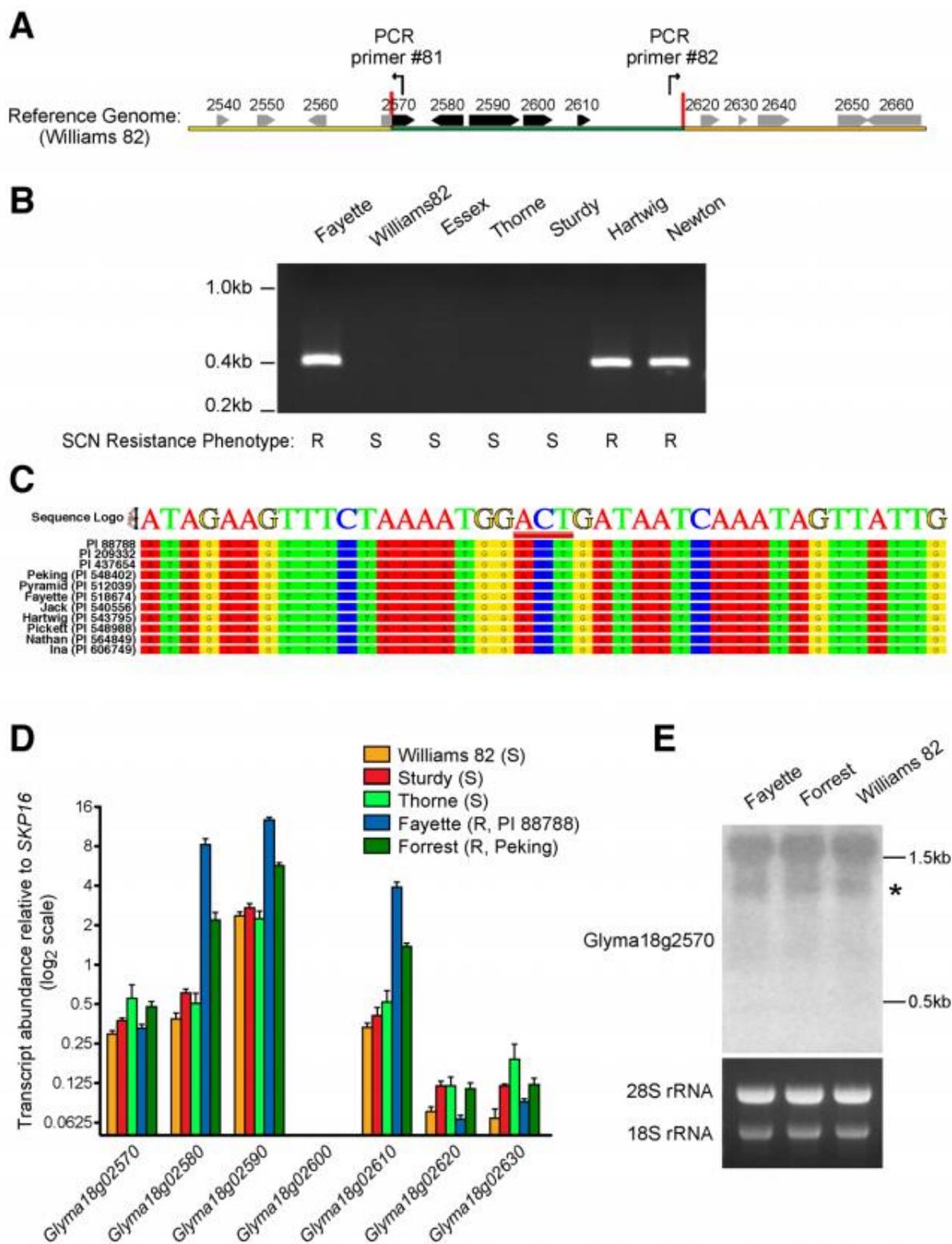
**Fig. 4.** Elevated SCN resistance conferred by simultaneous overexpression of multiple genes rather than overexpression of individual genes from the 31 kb *rhg1-b* repeat. SCN development beyond J2 stage is reported for transgenic soybean roots (variety Williams 82) overexpressing the designated single 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  $\pm$  std. error of mean for roots transformed with empty vector (EV) or gene overexpression constructs (OX). \*: Williams 82 - OX significantly different from Williams 82 - EV based on ANOVA  $p < 0.05$ .

## 2.6 Supplemental Figures



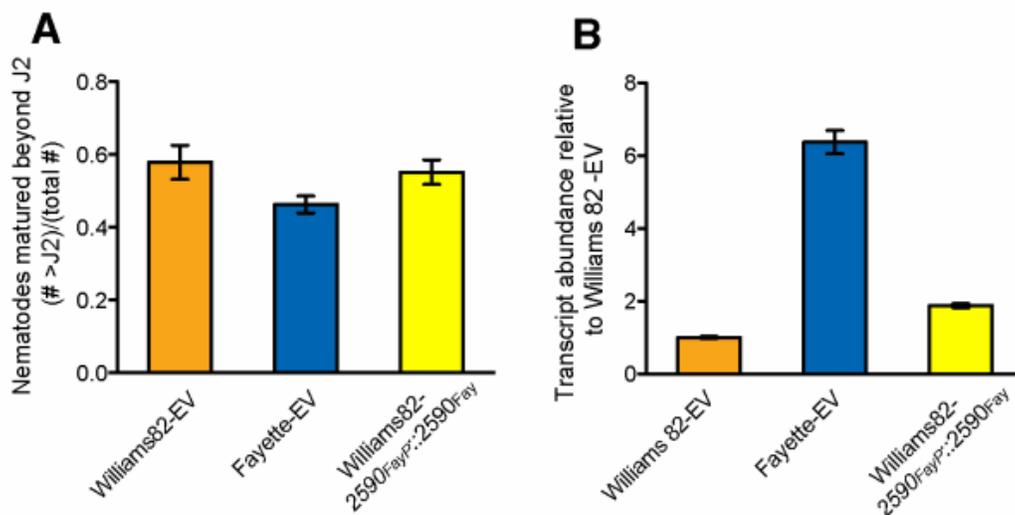
**Figure S1.** Nematode development is impacted by level of silencing. (A, C) Nematode development on Williams 82 and Fayette roots transformed with empty vector (EV), or Fayette transformed with silencing constructs (2580RNAi or ami2590) was dependent on level of silencing. Transgenic roots with reduced 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)** *Glyma18g02610*. Bars represent mean  $\pm$  std. error of mean.

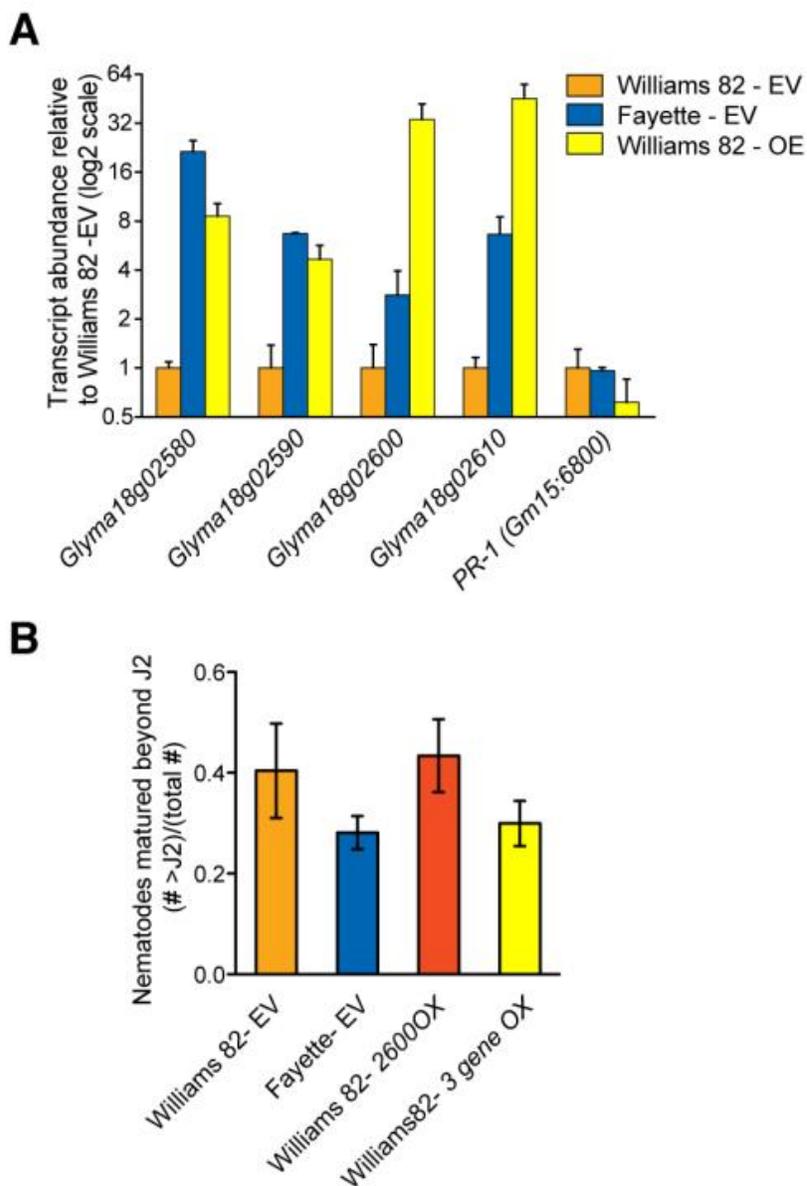


**Figure S2.** Multiple SCN-resistant varieties contain the DNA junction indicative of a repeat within the *Rhg1* locus, and exhibit elevated expression of genes fully encoded within the repeat.

(A) Schematic of PCR primers used in (B) (see also Figure 2). (B) Results of PCR using 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). (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  $\pm$  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  $\pm$  std. error of mean.



**Figure S4.** 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. 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  $\pm$  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).

## 2.7 Supplemental Tables

**Table S1.** Length estimates for Fiber-FISH hybridization signals. N: number of DNA fibers analyzed.

	Length of Fiber-FISH signals ( $\mu\text{m}$ )	Estimated Length of Fiber-FISH signals (kb)	N
Williams 82	$8.60 \pm 0.49$	$27.60 \pm 1.58$	20
Fayette	$96.56 \pm 7.54$	$309.94 \pm 24.2$	20
Peking	$34.42 \pm 2.91$	$110.48 \pm 9.34$	20

**Table S1.**

**Table S2.** Amino acid polymorphisms within the 31 kb repeat at *rhg1-b*. The position of the protein-coding genes was predicted by comparing the fosmid clone sequences to the Glymal 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 (Glymal; [www.phytozome.net](http://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 S2.**

**Table S3.** DNA sequences of oligonucleotide primers used for PCR (3 pages).

<b>ID</b>	<b>Primer</b>	<b>Sequence</b>
<b>Silencing constructs</b>		
<b>1</b>	2570 hpRNAi _F	AGGATCCATTTAAATCAAGTACTCTTCCCCACAAAAGCT
<b>2</b>	2570 hpRNAi _R	ACCTAGGAGGCGCGCCTGGGGCATTTCAGTAATTAGGTC
<b>3</b>	2580 hpRNAi _F	acctaggaggcgcgccTCATGAAGGTTCTCGGCGTAG
<b>4</b>	2580 hpRNAi _R	aggatccatttaaatCCACCAGTGAATTCCAAACCA
<b>5</b>	2590 hpRNAi _F	GAcctaggcgcgccGGACTTGGTCTGCAACACAGTC
<b>6</b>	2590 hpRNAi _R	GCg gatccatttaaatGAGCAGCAAAGTGGGCAACT
<b>7</b>	2600 hpRNAi _F	acctaggaggcgcgccGCCAAATTCAAAAGGCTTGCT
<b>8</b>	2600 hpRNAi _R	aggatccatttaaatCACCATTCAACATGCCTGTCA
<b>9</b>	2610 hpRNAi _F	taacctaggaggcgcgccACAACCTCTTCCGATTCTGTTCCG
<b>10</b>	2610 hpRNAi _R	caggatccatttaaatAGATACAACCACCTGAATACGCC
<b>11</b>	2620 hpRNAi _F	AGGATCCATTTAAATCTCGCAACACCATATCCAGAGTA
<b>12</b>	2620 hpRNAi _R	ACCTAGGAGGCGCGCCGGTGTAAAGGTCGAACCTGCGAA
<b>13</b>	2590-1 I miR-s	gaTATTGGTTATAGCAACACCGTtctctctttgtattcc
<b>14</b>	2590-1 II miR-a	gaACTTTGCTATAACCAATAtcaaagagaatcaatga
<b>15</b>	2590-1 III miR*s	gaACAGTGTGCTATTACCAATTtcacaggtcgtgatatg
<b>16</b>	2590-1 IV miR*a	gaAATTGGTAATAGCAACACTGTtctacatatattctct
<b>17</b>	2610-1 I miR-s	gaTATTTCCCGACCCGACGGGACTctctctttgtattcc
<b>18</b>	2610-1 II miR-a	gaTCCCCGTCGGGTCGGGAAATAtcaaagagaatcaatga
<b>19</b>	2610-1 III miR*s	gaGTACCGTCGGGTCGGGAAATtcacaggtcgtgatatg
<b>20</b>	2610-1 IV miR*a	gaAATTTCCCGACCCGACGGTACTctacatatattctct
<b>21</b>	2610-2 I miR-s	gaTATCCAGTCACCGGACGTGGtctctctttgtattcc
<b>22</b>	2610-2 II miR-a	gaCCACGTCGCGGTGACTGGATAtcaaagagaatcaatga
<b>23</b>	2610-2 III miR*s	gaCCCCGTCGCGGTGTCTGGATTtcacaggtcgtgatatg
<b>24</b>	2610-2 IV miR*a	gaAATCCAGACACCGGACGGGtctacatatattctct
<b>Transcript Abundance using qPCR</b>		
<b>25</b>	gm18: 2570 F_qPCR	TGAGATGGGTGGAGCTCAAGAAC
<b>26</b>	gm18: 2570 R_qPCR	AGCTTCATCTGATTGTGACAGTGC
<b>27</b>	gm18: 2580 F_qPCR	CGTGTAGAGTCCTTGAAGTACAGC
<b>28</b>	gm18: 2580 R_qPCR	ACCAGAGCTGTGATAGCCAACC
<b>29</b>	gm18: 2590 F_qPCR	TCGCCAAATCATGGGACAAGGC
<b>30</b>	gm18: 2590 R_qPCR	CAATGTGCAGCATCGACATGGG
<b>31</b>	gm18: 2600 F_qPCR	GCTTCAGTCAAGAAAATGTGCATG
<b>32</b>	gm18: 2600 R_qPCR	CACCCGAAACCGCGACACAAATG
<b>33</b>	gm18: 2610 F_qPCR	AGGTCACGTGTTGCCGTTG
<b>34</b>	gm18: 2610 R_qPCR	AAACCACACCAATAACAACAAAGCTCT
<b>35</b>	gm18: 2620 F_qPCR	AAGCCCAACAGGCCAAAGAGAG
<b>36</b>	gm18: 2620 R_qPCR	ACACCAAATGGGTTTCGCACTTC
<b>37</b>	gm18: 2630 F_qPCR	TTGTGGAAGTGAAAGTCGGTTTGC
<b>38</b>	gm18: 2630 R_qPCR	GTTGTACGTTTTCCCGTAACAATG
<b>39</b>	EF1b_For.qRT	CCACTGCTGAAGAAGATGATGATG
<b>40</b>	EF1b_Rev.qRT	AAGGACAGAAGACTTGCCACTC
<b>41</b>	SKIP16_For.qRT	GAGCCCAAGACATTGCGAGAG
<b>42</b>	SKIP16_Rev.qRT	CGGAAGCGGAAGAACTGAACC

<b>43</b>	UKN2_For.qRT	GCCTCTGGATACCTGCTCAAG
<b>44</b>	UKN2_Rev.qRT	ACCTCCTCCTCAAACCTCCTCTG
<b>45</b>	ACT11_For.qPCR	ATCTTGACTGAGCGTGGTTATTCC
<b>46</b>	ACT11_Rev.qPCR	GCTGGTCCTGGCTGTCTCC
<b>47</b>	UNK1_For.qPCR	TGGTGCTGCCGCTATTTACTG
<b>48</b>	UNK1_Rev.qPCR	GGTGAAGGAACTGCTAACAATC
<b>49</b>	TIP41_For.qPCR	AGGATGAACTCGCTGATAATGG
<b>50</b>	TIP41_Rev.qPCR	CAGAAACGCAACAGAAGAAACC
<b>51</b>	PR-1 (6790) F	TGCTTGGTCACTGGAAGTTGG
<b>52</b>	PR-1 (6790) R	AACTTCCTGCGAGCTGCGATAC
<b>53</b>	PR-1 (6800) F	AGTCATTGTGGGTGATCATGCTG
<b>54</b>	PR-1 (6800) R	GCAGCGTTGTGTGCATTAACAAAG

#### Expression Vectors

<b>55</b>	Ox2610-SalF2	gtcgacATGCGCATGCTCACCGG
<b>56</b>	P2610fused-R	TATTGCGAGAACCAAACCGG
<b>57</b>	Ox2590Sal-F	GGgtcgacATGGCCGATCAGTTATCGAAGG
<b>58</b>	Ox2590fused-R	AGTAATAGCCTCATGCTGCTCAAGTT
<b>59</b>	TerXba-R	ACtctagaGCGCATGTCTTTCGTTGATG
<b>60</b>	GmubiXba-F	GCTctagaGGGCCCAATATAACAACGACG
<b>61</b>	TerKpn-R	TCggtaccGCGCATGTCTTTCGTTGATG
<b>62</b>	PPA Linker_Top	GatgtcTTAATTAAtatctgtGGGCCcactatGGCGCGCCaatgtaaA
<b>63</b>	PPA Linker_Bottom	AGCTTttacattGGCGCGCCatagtGGGCCCacagataTTAATTAAgacatCTGCA
<b>64</b>	Ox2600-F	ATGGTTTCGGTTGATGATGGG
<b>65</b>	Ox2600-R	TTTTTGTGCATATAAGGGGTTTCAT
<b>66</b>	NosHind-F	GCaagcttGATCATGAGCGGAGAATTAAGGG
<b>67</b>	Nos2600-R	CCCATCATCAACCGAAACCATAGATCCGGTGCAGATTATTTGG
<b>68</b>	Nos2600-F	CCAAATAATCTGCACCGGATCTATGGTTTCGGTTGATGATGGG
<b>69</b>	NosAsc-R	TCggcgcgccGCGCATGTCTTTCGTTGATG
<b>70</b>	Ox2580-F	ATGTCTCCGGCCGCCG
<b>71</b>	Ox2580-R	TGACTTGCTACTAAAAGCATTATATATGTTG
<b>72</b>	NosAsc-F	CAggcgcgccGATCATGAGCGGAGAATTAAGGG
<b>73</b>	Nos2580-R	CGGCGGCCGGAGACATAGATCCGGTGCAGATTATTTGG
<b>74</b>	Nos2580-F	CCAAATAATCTGCACCGGATCTATGTCTCCGGCCGCCG
<b>75</b>	NosSbf-R	TGcctgcaggGCGCATGTCTTTCGTTGATG
<b>76</b>	M13 F	GTAAAACGACGGCCAG
<b>77</b>	M13 R	CAGGAAACAGCTATGAC
<b>78</b>	pSM101 seq	GTCTTGATGAGACCTGCTGCG
<b>79</b>	g2590pHind-F	CTTaagcttGAATGGTTTTTGTGTTTGTGTCTCTCAC
<b>80</b>	g2590pSal-R	TTGGTCGACCGTATCATCCAATG

#### Bridge PCR

<b>81</b>	SCN_Res Bridge F	TTTAGCCTGCTCCTCACAAATTC
<b>82</b>	SCN_Res Bridge R	TTGGAGAATATGCTCTCGGTTGT

#### Probes for Northern Analysis and Alt Transcript

<b>83</b>	2570F_qPCR	TGAGATGGGTGGAGCTCAAGAAC
<b>84</b>	2570 UTR Rev	CAAGTACTCTTCCCCACAAAAGC
<b>85</b>	2570 put exon F	TGCAGTTTTAGTGAAAGGCC

<b>86</b>	2570 exon 6 R	TCATCAAGCTCAACTTGAATCCC
<b>RACE PCR</b>		
<b>87</b>	2590-5GSP	GATCGGCCATTTTCTCCGATCGAAACA
<b>88</b>	2590-5NGSP	GACGACCAAGTCCAAATCCAAAACCCGC
<b>89</b>	2590-3GSP	AAGCCAAAGAACTTGAGCAGCATGAGGC
<b>90</b>	2590-3NGSP	CTGTCCAGTTGTTTCGTCTTACACATCCA
<b>91</b>	2610-5GSP	GGCGACGATCTTGACGACGGCGTT
<b>92</b>	2610-5NGSP	TCATACAGTGCAACCACCAGCCGCG
<b>93</b>	2610-3GSP	GGACGAGGTCACGTGTTGCCGTTGCT
<b>94</b>	2610-3NGSP	TTCACCACTATGGGCGTATTCAGGTGGT
<b>95</b>	2580-3GSP	CCTGGGGGATTCCAAAGGAACGC

---

**Table S3.**

## 2.8 References

- 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
- Bennetzen JL** (2000) Transposable element contributions to plant gene and genome evolution. *Plant Molecular Biology* **42**: 251-269
- Bybd DW, Kirkpatrick T, Barker KR** (1983) AN IMPROVED TECHNIQUE FOR CLEARING AND STAINING PLANT-TISSUES FOR DETECTION OF NEMATODES. *Journal of Nematology* **15**: 142-143
- 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
- Concibido VC, Diers BW, Arelli PR** (2004) A decade of QTL mapping for cyst nematode resistance in soybean. *Crop Science* **44**: 1121-1131
- 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
- Davis EL, Hussey RS, Mitchum MG, Baum TJ** (2008) Parasitism proteins in nematode-plant interactions. *Current Opinion in Plant Biology* **11**: 360-366
- Dodds PN, Rathjen JP** (2010) Plant immunity: towards an integrated view of plant-pathogen interactions. *Nat Rev Genet* **11**: 539-548
- Doyle JJ, Dickson EE** (1987) PRESERVATION OF PLANT-SAMPLES FOR DNA RESTRICTION ENDONUCLEASE ANALYSIS. *Taxon* **36**: 715-722
- Field B, Fiston-Lavier AS, Kemen A, Geisler K, Quesneville H, Osbourn AE** (2011) Formation of plant metabolic gene clusters within dynamic chromosomal regions. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 16116-16121
- Frei dit Frey N, Robatzek S** (2009) Trafficking vesicles: pro or contra pathogens? *Current Opinion in Plant Biology* **12**: 437-443
- Gheysen G, Mitchum MG** (2011) How nematodes manipulate plant development pathways for infection. *Current Opinion in Plant Biology* **14**: 415-421

- 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
- 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
- Huang GZ, Allen R, Davis EL, Baum TJ, Hussey RS** (2006) Engineering broad root-knot resistance in transgenic plants by RNAi silencing of a conserved and essential root-knot nematode parasitism gene. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 14302-14306
- 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
- Kelley LA, Sternberg MJE** (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols* **4**: 363-371
- 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 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
- M.M.** (2012) Materials and methods are available as supplementary materials on Science Online.
- 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*
- McHale LK, Haun WJ, Xu WW, Bhaskar PB, Anderson JE, Hyten DL, Gerhardt DJ, Jeddloh JA, Stupar RM** (2012) Structural variants in the soybean genome localize to clusters of biotic stress response genes. *Plant Physiol*
- 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**

- 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
- Niblack TL, Lambert KN, Tylka GL** (2006) A model plant pathogen from the kingdom animalia: Heterodera glycines, the soybean cyst nematode. *In Annual Review of Phytopathology*, Vol 44, pp 283-303
- Okumoto S, Pilot G** (2011) Amino Acid Export in Plants: A Missing Link in Nitrogen Cycling. *Molecular Plant* **4**: 453-463
- Ossowski S, Schwab R, Weigel D** (2008) Gene silencing in plants using artificial microRNAs and other small RNAs. *Plant J* **53**: 674-690
- 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
- 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
- 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
- Sambrook J, Russell DW** (2001) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York
- 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
- 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
- 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
- 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
- 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**
- Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I** (2009) ABySS: A parallel assembler for short read sequence data. *Genome Research* **19**: 1117-1123
- Smith NA, Singh SP, Wang MB, Stoutjesdijk PA, Green AG, Waterhouse PM** (2000) Total silencing by intron-spliced hairpin RNAs. *Nature* **407**: 319-320
- Springer NM, Ying K, Fu Y, Ji TM, Yeh CT, Jia Y, Wu W, Richmond T, Kitzman J, Rosenbaum H, Iniguez AL, Barbazuk WB, Jeddloh 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. *In Annual Review of Medicine*, Vol 61. Annual Reviews, Palo Alto, pp 437-455
- 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
- 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
- Winzer T, Gazda V, He Z, Kaminski F, Kern M, Larson TR, Li Y, Meade F, Teodor R, Vaistij FE, Walker C, Bowser TA, Graham IA** (2012) A Papaver somniferum 10-Gene Cluster for Synthesis of the Anticancer Alkaloid Noscapine. *Science* **336**: 1704-1708
- 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
- 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
- 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
- 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

### **Chapter 3: Distinct copy number, coding sequence and locus methylation patterns underlie *Rhg1*-mediated soybean resistance to soybean cyst nematode**

This chapter was previously published as:

Cook DE, Bayless AM, Wang K, Guo X, Song Q, Jiang J, Bent AF (2014) Distinct Copy Number, Coding Sequence, and Locus Methylation Patterns Underlie *Rhg1*-Mediated Soybean Resistance to Soybean Cyst Nematode. *Plant Physiol* 165: 630-647

Contributions: David Cook wrote the manuscript and contributed the majority of the described work. I performed copy number qPCR in Fig. 1A, assisted Kai Wang's Fiber-FISH analysis in Fig. 2B, contributed Fig. 3 (identified and cloned the *Rhg1* low copy type  $\alpha$ -SNAPs), and made occasional conceptual contributions throughout the project, and made minor contributions to manuscript preparation.

### 3.1 Abstract

Copy number variation of kilobase-scale genomic DNA segments, beyond presence/absence polymorphisms, can be an important driver of adaptive traits. *Rhg1* is a widely utilized quantitative trait locus that makes the strongest known contribution to resistance against soybean cyst nematode (SCN; *Heterodera glycines*), the most damaging pathogen of soybean (*Glycine max*). *Rhg1* was recently discovered to be a complex locus at which resistance-conferring haplotypes carry up to ten tandem repeat copies of a 31 kb DNA segment, and three disparate genes present on each repeat contribute to SCN resistance. Here we use whole-genome sequencing, fiber-FISH and other methods to discover the genetic variation at *Rhg1* across 41 diverse soybean accessions. Based on copy number variation, transcript abundance, nucleic acid polymorphisms and differentially methylated DNA regions, we find that SCN resistance is associated with multi-copy *Rhg1* haplotypes that form two distinct groups. The tested high copy-number *Rhg1* accessions, including PI 88788, contain a flexible number of copies (7 to 10) of the 31 kb *Rhg1* repeat. The identified low copy-number *Rhg1* group, including Peking and PI 437654, contains 3 copies of the *Rhg1* repeat and a newly identified allele of *Glyma18g02590* (predicted  $\alpha$ -SNAP). There is strong evidence for a shared origin of the two resistance-conferring multi-copy *Rhg1* groups and subsequent independent evolution. Differentially methylated DNA regions also were identified within *Rhg1*, that correlate with soybean cyst nematode resistance. The data provide insights into copy number variation of multi-gene segments, using as the example a disease resistance trait of high economic importance.

### 3.2 Introduction

Vascular plants experienced a rapid diversification following land colonization, overcoming biotic and abiotic stresses to occupy diverse niches in a process that continues to the present and includes human-guided plant breeding (Kenrick and Crane, 1997; Steemans et al., 2009; Oh et al., 2012). One mechanism of genetic variation is diversification of the physical genome, at scales broader than isolated DNA base-pair changes. This genome structural variation (Feuk et al., 2006) is increasingly recognized for having significant impacts on phenotypes and evolution (Aitman et al., 2006; Perry et al., 2008; Maron et al., 2013). Recent advances in plant genomics have highlighted the role of structural variation in plant adaptation to environmental stress (DeBolt, 2010; Dassanayake et al., 2011; Wu et al., 2012; Olsen and Wendel, 2013).

Copy number variation is an important type of structural variation because of its varied evolutionary impacts, facilitating 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 are not retained, undergo pseudogenation, or exhibit distinct negative effects (Lynch and Conery, 2000; Demuth and Hahn, 2009; Tang and Amon, 2013), gene duplication has facilitated evolution in diverse organisms (Kondrashov et al., 2002; Conant and Wolfe, 2008). For one of the simplest types of copy number variation, gene duplication, a wide range of resulting adaptations to changing local environmental conditions have been characterized (Triglia et al., 1991; Labbe et al., 2007; Schmidt et al., 2010; Dassanayake et al., 2011; Heinberg et al., 2013), reviewed in (Kondrashov, 2012). Single gene copy number amplification has also been observed as an adaptive response to selective pressures (Bass and Field, 2011).

Epigenetic modifications, prominently including differential cytosine methylation, can also significantly impact organismal phenotypes (Chen, 2007; Gohlke et al., 2013; Herrando-Herraez et al., 2013). While the term epigenetic indicates heritable changes in gene activity not caused by changes in DNA sequence, there is increasing appreciation both of the extent of methylation and other epigenetic marks throughout genomes, and of the plasticity of these marks (Schmitz et al., 2013; Ziller et al., 2013).

Domesticated soybean (*Glycine max*) is an important world commodity, accounting for a majority of the world's protein-meal and oilseed production (soystats.com). The most economically damaging pathogen of soybean is the soybean cyst nematode (SCN), *Heterodera glycines* (Niblack et al., 2006). Soybean cyst nematodes are obligate endoparasites that cause disease by reprogramming host root cells to form specialized feeding cells termed syncytia, robbing the plant of carbon and adversely affecting yield (Lauritis et al., 1983; Endo, 1984; Young, 1996; Sharma, 1998). SCN is found in all major soybean-growing states in the US and cannot feasibly be removed (Niblack, 2005). Because the primary control strategies for SCN are crop rotation and planting resistant varieties, significant attention has been focused on the identification, development and use of soybean germplasm that exhibits resistance to SCN (Diers et al., 1997; Concibido et al., 2004; Brucker et al., 2005; Wrather and Koenning, 2009; Kim et al., 2010; Kim et al., 2011). The *Rhg1* (resistance to *Heterodera glycines*) locus, sometimes in combination with *Rhg4*, makes the greatest contribution to resistance in the vast majority of the commercially utilized soybean cultivars that exhibit SCN resistance (Caldwell et al., 1960; Matson and Williams, 1965; Webb et al., 1995; Li et al., 2004; Brucker et al., 2005; Tylka et al., 2012).

We recently discovered that the SCN resistance conferred by *Rhg1* is mediated by a 31 kb segment of DNA that contains four open reading frames and exhibits substantial copy number variation (Cook et al., 2012). A commercial soybean line containing the most widely utilized version of the *Rhg1* locus, derived from plant introduction (PI) 88788, contains 10 tandem repeat copies of the 31 kb segment. Only a single copy of this 31 kb block was detected in the SCN-susceptible line Williams 82 and three other SCN-susceptible lines. It is particularly intriguing that three distinct genes within the 31 kb repeat were shown to contribute to SCN resistance (Cook et al., 2012). These genes are *Glyma18g02580* (encoding a predicted amino acid transporter), *Glyma18g02590* (encoding a predicted  $\alpha$ -SNAP vesicle trafficking protein), and *Glyma18g02610* (encoding a protein lacking a predicted function). The predicted protein sequences of *Glyma18g02580* and *Glyma18g02610* were invariant between the examined SCN-resistant and SCN-susceptible alleles, and experimental evidence suggests that these two genes contribute to resistance via enhanced expression arising through copy number variation. The SCN-resistant line derived from PI 88788 did contain an alternate allele of *Glyma18g02590*, which was also more highly expressed in SCN-resistant lines relative to susceptible lines. In addition to PI 88788, 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 (Cook et al., 2012).

A well-documented epistasis occurs in Peking-derived SCN resistance, in which Peking *Rhg1* has low efficacy relative to the *Rhg1* from PI 88788, but only if Peking *Rhg4* is not simultaneously present (Brucker et al., 2005; Liu et al., 2012). The responsible gene at *Rhg4*

was recently discovered to encode a serine hydroxymethyltransferase (SHMT) (Liu et al., 2012). Peking and PI 437654 (the source of the less used but commercially relevant ‘Hartwig’ or ‘CystX’ resistance), contain an *Rhg4* allele whose product exhibits altered enzyme kinetics. Impacts of *Rhg4* on SCN resistance are difficult to detect when deployed together with the high copy-number *rhg1-b* from PI 88788 (Brucker et al., 2005). It is intriguing and of high economic relevance that SCN populations arise that partially overcome the resistance mediated by certain sources of *Rhg1* while remaining sensitive to the resistance conferred by other *Rhg1* sources (Niblack et al., 2002; Colgrove and Niblack, 2008). In addition to understanding the biology of trait variation caused by copy number variation, and of traits in multicellular eukaryotes that are conferred by tightly linked blocks of distinct genes, there is substantial practical interest in understanding the variation in SCN resistance caused by different sources of *Rhg1*, and in the potential to predict, discover and/or develop more effective versions of *Rhg1*.

Here we use qPCR, fiber-FISH, whole-genome sequencing, and DNA methylation analyses to investigate the major SCN resistance locus *Rhg1* from a diverse population of soybean lines. We sequenced and analyzed the genomes of six “Hg Type Test” soybean lines that are widely used to characterize SCN field populations for their capacity to overcome different sources of SCN resistance (Niblack et al., 2002), and also analyzed whole genome sequence data from 35 diverse soybean lines that are in use as parents in a separate SoyNAM (nested association mapping) project. We discovered three classes of the *Rhg1* locus that can be differentiated by gene dosage, copy number, and coding sequence. We also observed differential DNA methylation between resistant and susceptible *Rhg1* haplotypes, at genes impacting SCN resistance. The collective data allow clearer inferences to be drawn regarding the evolutionary history of the locus, and provide a detailed analysis of one of the few confirmed examples in

plant or animal biology in which copy number variation of a small multi-gene segment contributes to a defined adaptive trait.

### 3.3 Results

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

To assess the natural variation present at *Rhg1*, beyond the previous determination that there are ten and three copies of the 31 kb *Rhg1* repeat in two previously studied lines (Cook et al., 2012), we analyzed five other SCN-resistant lines. Together with PI 88788 and Peking, these seven soybean lines comprise the diagnostic test set in the established Hg Type Test that describes the capacity of SCN populations to overcome different sources of SCN resistance (Niblack et al., 2002). Initial characterization of *Rhg1* copy number, using qPCR on genomic DNA, revealed three copy number classes: single-copy, low-copy (2 to 4 copies) and high-copy (>6 copies) (Fig. 1A). For lines estimated to contain >6 copies, qPCR produced variable results and unreliable 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 constitutive transcription, we quantified root transcript abundance using qPCR in the Hg Type Test lines (Niblack et al., 2002). 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 Test SCN resistance lines, relative to SCN-susceptible Williams 82 (Fig. 1B). The transcript abundance of an adjacent gene that is outside of the 31 kb repeat, *Glyma18g02570*, had similar transcript abundance across all tested SCN-resistant and SCN-susceptible genotypes. Four of the SCN-resistant genotypes, Peking, PI 90763, PI 89772, and PI 437654, showed similar levels of elevated expression of the repeated

genes, while expression was even more elevated in Cloud (PI 548316), PI 88788, and PI 209332 (Fig. 1B). These groupings were the same as those identified for qPCR estimates of DNA copy number, and indicate that transcript abundance for these genes scales with gene copy number. One gene in the repeat, *Glyma18g02600*, was more highly expressed in SCN-resistant lines but the expression level was similar between genotypes in different copy number classes. However, transcript abundance for this gene was close to the limit of detection for qPCR, was also 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). The soybean line Cloud, which was placed in the high-copy number class but estimated to have fewer *Rhg1* copies than PI 88788 and PI 209332, also showed lower transcript abundance of *Glyma18g02580* and *Glyma18g02590* than the other two lines in the high-copy class (Fig. 1B).

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

To definitively determine *Rhg1* copy number in the Hg Type Test 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). Representative fiber-FISH images for soybean lines PI 90763, PI 89772, and PI 437654 shown in 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 head-to-tail direct repeats. These results confirm the copy number estimates from qPCR. More importantly, for soybean lines in the high-copy *Rhg1* class, fiber-FISH precisely determined the presence of seven *Rhg1* copies in Cloud, nine copies in PI 88788 and ten copies in PI 209332 (Fig. 2B, bottom 3 panels). We had previously used fiber-FISH to determine that Fayette, a soybean variety containing a *Rhg1* locus originally from PI 88788, carries ten copies

of the *Rhg1* repeat (Cook et al., 2012). Hence the number of *Rhg1* repeats varies not only between haplotype classes, but also within the high-copy class and between lines with recent shared ancestry.

### **Read depth analysis from whole genome sequencing identifies *Rhg1* copy number and predicts SCN resistance**

To further discover the nature of the diversity within soybean *Rhg1*, we performed whole genome sequencing for six of the seven Hg Type Test 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 shotgun sequence data for 35 diverse soybean lines, generated as part of the recently initiated SoyNAM project, and analyzed previously published Illumina sequencing data from an undomesticated *Glycine soja* accession (Kim et al.). Supplemental Table 1 and 2 provide details regarding the sequencing datasets. For the present study we focused on in-depth analysis of *Rhg1* on chromosome 18 and its paralogous locus on chromosome 11.

To initially uncover structural variation at *Rhg1*, we screened the SoyNAM genome sequence data sets by aligning Illumina reads to a portion of the Williams 82 reference genome corresponding to *Rhg1* on soybean chromosome 18 and similar loci (see Methods). This screen determined that 8 of 35 SoyNAM lines contain an estimated *Rhg1* copy number greater than 1, based on read depth across the known repeat and flanking regions (Supplemental Table S3). To further investigate the extent of copy number variation in this set of diverse soybean genomes and to eliminate possible mapping bias that might arise from use of a limited reference sequence region, Illumina sequencing reads were re-mapped to the entire reference genome for 24 of the

SoyNAM lines based on the results of the rapid alignment and sequencing depth. This provided more precise *Rhg1* copy number estimates based on read depth. Along with the SoyNAM lines, six Hg Type Test lines sequenced as part of this work, and the available *G. soja* genome sequence were included for in-depth analysis. As shown in Figure 2C, the estimated copy numbers based on read depth for the six Hg Type Test 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 (Fig. 2C). A soybean line derived from PI 88788 was previously estimated to carry ten copies of the *Rhg1* repeat, using read-depth analysis of whole genome sequence data (Cook et al., 2012). The majority of the soybean lines chosen for the SoyNAM study were found to carry a single copy of the *Rhg1* locus (Fig. 2C; Supplemental Table S3), and have not been reported to exhibit SCN resistance where information is publicly available from the Germplasm resources information network (GRIN) (USDA). Seven other SoyNAM lines contain 9 to 10 copies of the *Rhg1* locus, while one line contains an estimated 3 copies (Fig. 2C). These results are in agreement with pedigree information where it is publicly available. The above results indicate that increased copy number at *Rhg1* is not a common phenomenon in *Glycine max* accessions, and likely 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 (Fig. 2D).

**Sequence analysis reveals extensive *Rhg1* locus DNA sequence variation, but amino acid polymorphisms are only present in the predicted  $\alpha$ -SNAP.**

The whole genome sequence data of the SoyNAM and Hg Type Test lines were analyzed for *Rhg1* nucleic acid and derived amino acid variation. Genomic DNA sequence variations,

including single nucleotide polymorphisms (SNPs), and small insertions and deletions relative to the Williams 82 reference genome, 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 *Rhg1* repeat interval (chromosome 18 bp #1,632,223 – 1,663,500 of the Williams 82 soybean reference genome, version 1.1) were identified in at least one of the 31 genomes. The average number of *Rhg1* variant sites per soybean line was  $251 \pm 40$  (mean  $\pm$  SE) for the low-copy *Rhg1* lines, and was  $260 \pm 10$  for the high-copy lines, while it was  $23 \pm 29$  for the lines estimated to contain a single copy of *Rhg1* (see Table I for full results). As is further described in a later section, within any single accession the sequences of individual repeats were largely identical to the other repeats. Hence there were approximately 250 polymorphisms per 31 kb repeat in the SCN-resistant genotypes, but only zero to 81 polymorphisms in the corresponding 31 kb *Rhg1* region of the investigated single-copy lines.

Despite the high number of sequence polymorphisms found within each *Rhg1* repeat in SCN-resistant lines, few affected protein-coding sequences. We did not detect any polymorphisms resulting in an altered amino acid sequence for *Glyma18g02610* or *Glyma18g02580*, in any of the SCN-resistant lines. Curiously, in the derived amino acid sequences of *Glyma18g02590*, two SCN-resistant allele types were observed that carry distinct mutations, but which impact similar protein sites (Fig. 3A). The gene *Glyma18g02590* encodes a predicted  $\alpha$ -SNAP; in other organisms these proteins have the canonical function of stimulating NSF ATPase activity to assist the disassembly of SNARE components following vesicle-mediated transport (Morgan et al., 1994; Barnard et al., 1997; Rice and Brunger, 1999). Amino acid sequence alignment of the available 17 *Rhg1* single-copy soybean lines including the Williams 82 reference genome revealed an invariant primary sequence of *Glyma18g02590*. One

type of alternative allele was found in all tested high-copy *Rhg1* haplotypes, including the previously reported sequence from Fayette, and a new allele was found in all tested lines carrying the low-copy *Rhg1* haplotype associated with SCN resistance (Fig. 3B). The novel alleles of  $\alpha$ -SNAP found in SCN-resistant lines have amino acid polymorphisms changing the final 5 or 6 amino acids, the residues that otherwise have the strongest consensus sequence across eukaryote  $\alpha$ -SNAPs (Fig. 3C), including a substitution for the leucine at the penultimate C-terminal amino acid. The presence of different amino acid substitutions at similar positions between the low and high-copy class 2590 alleles suggests a functional importance of these sites for SCN disease resistance. Mutations at these C-terminal residues are unexpected given previous findings that these residues are essential for stimulating N-ethylmaleimide-sensitive factor (NSF) ATPase activity in other organisms (Barnard et al., 1996). Together, these findings suggest that the SCN resistance-associated *Glyma18g02590* proteins may not possess classical  $\alpha$ -SNAP functions, and may instead promote SCN disease resistance through a novel mechanism.

For *Glyma18g02590* we performed 3'RACE and sequenced at least 7 independent cDNA clones for each of the Hg Type Test lines and Williams 82. The novel (non-Williams 82) *Glyma18g02590* alleles predicted from genomic DNA sequences were present in cDNA from the respective lines carrying the low- or high-copy *Rhg1* haplotypes (Supplemental Table S4). Interestingly, a small proportion of the cDNA clones sequenced from PI 88788 and Cloud (high-copy *Rhg1* lines) contained Williams 82-type *Glyma18g02590* sequences, consistent with the identification of a single Williams 82-type genomic DNA sequence in one of the copies of the 31 kb *Rhg1* repeats (described below). We did not detect any Williams 82-type *Glyma18g02590* sequences in cDNAs from lines carrying the low-copy class *Rhg1*, again consistent with the

genomic DNA sequence data. However, a splice-isoform of the *Glyma18g02590* cDNA was identified in all of the tested low-copy *Rhg1* lines, and this splice-isoform was not found in the high-copy or single-copy *Rhg1* lines (Supplemental Table S4; Fig. 3B).

A naturally occurring truncated allele encoding a predicted  $\alpha$ -SNAP was recently implicated in SCN disease resistance derived from Peking and PI 437654, but not PI 88788-derived resistance (Matsye et al., 2012). Our results from whole genome sequencing indicate, however, that the sequence encoding that truncated  $\alpha$ -SNAP is not encoded by a *Glyma18g02590* gene at *Rhg1* on chromosome 18, but rather by *Glyma11g35820* - the paralog of *Glyma18g02590* on chromosome 11 (Supplemental Fig. S1; Supplemental Table S5). The SNPs at *Glyma11g35820* responsible for encoding the truncated allele were also identified in the high-copy *Rhg1* SCN-resistant lines Cloud and LG05-4292.

Another *Rhg1* sequence polymorphism was identified in the Peking genome: a nucleotide deletion in the second exon of the *Glyma18g02600* coding sequence (Table II), observed as a heterozygous deletion (see below). Translation of the resulting mRNA results in a stop codon eight codons downstream of the deletion, truncating the predicted protein by 314 amino acids (removing 58% of the wild-type protein sequence).

### **Resistance-conferring *Rhg1* loci developed from a common source but underwent copy number expansion in distinct lineages**

To further explore the evolutionary history of the *Rhg1* locus, DNA sequence variation sites in a diverse set of soybean lines were used to construct a non-hierarchical phylogenetic network using the NeighborNet algorithm in Splits-tree (Bryant and Moulton, 2004; Huson and Bryant, 2006). The network reveals a clear split between the *Rhg1* loci from SCN-resistant

(right) and SCN-susceptible (left) lines (Fig. 4A). There is a further split in the multi-copy clade, separating the low- and high-copy *Rhg1* groups from each other (Fig. 4A). A common origin of the high-copy and low-copy *Rhg1* repeats was suggested by the identity of their repeat-junction sequences (Cook et al., 2012), and is now further supported by the high number of DNA sequence variant sites shared by the two groups but absent in single-copy *Rhg1* lines (Fig. 4B). In total, 147 DNA variant sites not detected in the single-copy *Rhg1* SCN-susceptible lines are common to all of the sequenced high-copy and low-copy Hg Type Test lines. This is 75% of the 197 DNA variant sites present in at least one Hg Type Test line but not present in any of the examined SCN-susceptible lines. The data suggest that a common progenitor had accumulated the 147 DNA variant sites prior to subsequent divergence of the two copy number groups. In support of subsequent divergence of the low-copy lines from the high-copy lines, a small number of DNA variant sites not present in any tested SCN-susceptible genome were universally common within either the low-copy *or* the high-copy *Rhg1* groups: 10 sites for low-copy and 7 for high-copy (Supplemental Fig. S2). Even more recent divergence is highlighted by the presence of a small number of DNA variants unique to a single tested genotype: Peking (6), PI 88788 (0), PI 90763 (1), PI 437654 (0), PI 209332 (5), PI 89772 (0), and Cloud (1).

The degrees of similarity between *Rhg1* repeats within any single genome or within a copy number group can be analyzed by the frequency of variant sequence relative to reference sequence, from the whole genome sequence data sets. Within the high-copy genomes of Cloud, PI 209332, and LD00-3309 (PI 88788 derivative), most of the variant sites on the right three-quarters of the interval as shown in Figure 5 have a sequence frequency of roughly 0.85- 0.9 (Fig. 5A). The other 10-15% of sequence reads at these positions match the Williams 82 reference sequence, suggesting that roughly three-quarters of one of the *Rhg1* repeats in the high-

copy *Rhg1* accessions contains Williams 82-type sequence. This is consistent with the *Glyma18g02590* cDNA data described above (Supplemental Table S4). Most variant sites across the left one-quarter of the *Rhg1* repeat (Fig. 5A) are invariant for the alternate sequence, indicating its presence in all copies.

A small number of DNA variant sites do not follow the above trend, and indicate the development and propagation of variant sequences in a smaller number of the total copies. Specifically, the DNA variant at chromosome 18 base-pair position 1657025 is apparently in only 4 of the 7 copies in Cloud, and in only 5 or 6 of the 10 copies in PI 209332 and LD00-3309, suggesting as one possibility the emergence of this DNA polymorphism in one repeat at an intermediate stage of copy number expansion of the locus (Supplemental Table S6). However, propagation of the repeats apparently was not symmetric between genomes, because (for example) at positions 1657807/1657816 and 1661264/1661293, Cloud and PI 209332 appear to carry only 5 and 7-8 copies respectively of the variant site while LD00-3309 appears to carry the variant site in all 9 non-Williams 82 repeats. Conversely, the set of polymorphisms at positions 1663007-1663250 are present in only 6-7 copies in LD00-3309, 8-9 copies in PI 209332, but are present in all 6 non-Williams 82 copies of Cloud (Supplemental Table S6). Inspection of raw sequence data for these non-homogeneous variant sites suggests that they are valid sequence calls rather than data processing errors, and suggests unequal propagation of specific copies during evolution of the locus. Although we cannot rule out phenotypic selection among the high-copy *Rhg1* soybean lines for revertants that carry more copies of the Williams 82 reference sequence at these non-homogeneous variant sites, the sites are in intergenic regions at least one kb away from known transcription start sites. Hence it may be more parsimonious to assume that

they are neutral sites that reflect the source of progenitor repeats that were utilized during *Rhg1* repeat expansion.

Analysis of the low-copy *Rhg1* lines (Peking, PI 89772, PI 90763, and PI 437654) shows a different pattern of repeat expansion and may partially account for well-established functional differences between the high-copy (PI 88788-type) and low-copy (Peking-type) *Rhg1* loci. The frequency of variant sequence to reference sequence at polymorphic sites in all *Rhg1* low-copy lines is nearly 1, i.e. mostly uniform across the 31.2 kb repeat region (Fig. 5B). This suggests that the low-copy lines experienced copy number expansion from a single shared progenitor, and/or, homogenization across the repeats by gene conversion or other mechanisms after at least some repeats had already formed. Loss of repeats carrying divergent copies may have also occurred. This in-depth analysis of sequencing frequencies shows that not only are the two resistance groups diverging for *Rhg1* copy number, the sequence composition of the repeats is also following different evolutionary paths.

### **Variation in soybean resistance to diverse nematode populations supports the high-copy and low-copy *Rhg1* groupings and suggests a relationship between copy number and resistance**

Previous research has described differences in SCN resistance between Peking-, PI 437654- and PI 88788-derived soybean sources, measured in terms of genetics, cell biology, nematode development, and nematode race-specificity or Hg Type-specificity, but the causes for these observations have remained elusive (Arelli and Webb, 1996; Mahalingam and Skorupska, 1996; Kim et al., 1998; Brucker et al., 2005; Niblack et al., 2006; Kim et al., 2010; Klink et al., 2011). To address this we analyzed data for soybean resistance to soybean cyst nematode, from

greenhouse trials conducted by Alison Colgrove, Terry Niblack and colleagues as part of the Northern Regional Soybean Trial (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 2009 to 2012, including SCN field populations from 8-10 north central U.S. states and/or adjacent Canada provinces per year. The results from our analysis indicate that Cloud, which contains 7 copies of *Rhg1*, was significantly less resistant than the other lines tested (Fig. 6). The other two lines in the high-copy *Rhg1* class, PI 88788 and PI 209332, which contain 9 and 10 copies respectively, form a statistically significantly more resistant cluster than Cloud, suggesting that higher *Rhg1* copy number may increase SCN resistance. Because Cloud, PI 88788 and PI 209332 lines are not isogenic at other loci, this comparison is only suggestive. The low-copy *Rhg1* lines are significantly more resistant to diverse SCN populations, but since these lines carry an SCN resistance-conferring allele of *Rhg4*, a simple comparison of the relative contributions or efficacies of *Rhg1* loci between low-copy and high-copy lines is obfuscated. Moreover, the role of low-copy and high-copy *Glyma18g02590* amino acid polymorphisms in impacting resistance to SCN is unknown.

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

In addition to determining the genome structure 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 root DNA methylation patterns at *Rhg1*, we used DNA methylation-sensitive restriction enzymes coupled with PCR to identify differentially methylated regions (DMR) between SCN-resistant and SCN-susceptible genotypes. The enzyme McrBC restricts

DNA at sites of methylated cytosines of the sequence (G/A)mC and does not restrict unmethylated DNA (Sutherland et al., 1992). Hence genomic DNA digestion by McrBC followed by PCR will not produce a product if the PCR product spans methylated cytosines. Using a total of 23 primer pairs, we discovered 8 DMRs between SCN-susceptible genomes (carrying a single-copy *Rhg1* locus) and SCN-resistant genomes (carrying low- or high-copy *Rhg1* loci) (Fig. 7). Hypermethylated DMRs were detected in SCN-resistant lines in the shared promoter for genes *Glyma18g02580* and *Glyma18g02590*, and within and flanking the coding sequence of *Glyma18g02610*. We did not observe DMRs in the gene body of *Glyma18g02580*, nor did we observe substantial methylation or DMRs adjacent to or within the coding sequence of *Glyma18g02600*. We also used McrBC to analyze methylation at the *Rhg1*-adjacent but non-repeated genes *Glyma18g02570* and *Glyma18g02620* and did not observe DMRs (Fig. 7).

During preparation of this manuscript, a genome-wide methylome study was published in which whole genome bisulfite sequencing was performed for soybean 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 resistance derived from PI 437654 (low-copy *Rhg1* locus type), while LDX contains a single copy of *Rhg1* (Diers et al., 2010; Kim et al., 2011). To confirm our observations and gain single-base resolution for methylation, we highlighted and re-analyzed the Schmitz *et al.* data, focusing on *Rhg1*.

Consistent with the findings described above, our *Rhg1* copy number estimate was 2.93 for LD and 1.17 for LDX, with various LD x LDX F3-derived (and hence potentially heterozygous) progeny families giving a range of *Rhg1* copy number estimates between 1 and 3 (Supplemental Fig. S3A). We were also able to estimate transcript abundance for the two parents along with the two F3-derived progeny families that were subjected to RNA-seq characterization

(see Methods). Consistent with our present and previous findings (Fig. 1B and (Cook et al., 2012)), standardized RNA sequence read depth for non-infested plants, normalized to the susceptible LDX parent, showed elevated expression for the genes encoded within but not adjacent to the *Rhg1* repeat in LD and progeny 11272, but not progeny 11268 (Supplemental Fig. S3B). This is consistent with elevated *Rhg1* copy number as a significant cause of the elevated transcript levels.

DNA methylation levels were computed from the Schmitz et al. data in bins of 150 bp in the CG, CHG and CHH sequence context, in both parents and 27 progeny lines that had at least a 4x average sequencing depth. Consistent with our above findings of differential root DNA methylation in different *Rhg1* copy-number groups, we observed differential hypermethylated DNA in all three-sequence contexts at the same regions in lines estimated to contain multiple copies of *Rhg1* (Fig. 8 and Supplemental Fig. S4). Data for the full set of lines can be seen in Figure S5 (see also methods). Consistent with the finding that methylation patterns are largely inherited based on the parental methylation pattern (Schmitz et al., 2013), for *Rhg1* we observed high average levels of cytosine methylation (a characteristic of the LD parent that carries three *Rhg1* copies) in the progeny that appeared homozygous for the three-copy *Rhg1* haplotype, and lower average *Rhg1* methylation (a characteristic of the LDX parent that carries one *Rhg1* copy) in the progeny homozygous for single-copy *Rhg1* haplotypes (Fig. 8B, 8D). Together, our findings and the data of Schmitz *et al.* describe in detail, across tissue types and different sources of SCN resistance, stably inherited hypermethylated DNA regions at the resistance-conferring alleles of the genes shown to mediate *Rhg1* resistance.

### 3.4 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 (Wrather and Koenning, 2009). Major efforts in soybean breeding and biotechnology are focused on the incorporation of desirable *Rhg1* alleles, and on continued discovery of new and better sources of SCN resistance. We had previously determined that three very tightly linked genes at *Rhg1* contribute to SCN resistance, and that these genes reside on a 31 kb segment that is present in ten copies in a common SCN-resistant variety along with an altered amino acid sequence for one of the genes (Cook et al., 2012). However, the extent of *Rhg1* structural variation present in a broader set of soybean germplasm, the presence of alternate coding alleles and their expression levels, and the relatedness of different *Rhg1* sources was not known. Here, we report the discovery of the structural, coding, and methylation differences present at *Rhg1* from a diverse population of soybean lines.

The identification in different soybean lines of 7, 9, and 10 copies of an *Rhg1* locus composed of highly similar sequences indicates that copy number at *Rhg1* is plastic, and malleable over the time scale of breeding cycles. This is evidenced by the discovery of 10 copies of *Rhg1* in Fayette, a line developed by backcrossing Williams 82 (single copy) to PI 88788 (9 copies) (Mikel et al., 2010). In contrast, all the sequenced SCN-resistant lines belonging to the low-copy *Rhg1* group contained 3 copies of nearly identical *Rhg1* repeats. It will be interesting to identify additional sources of SCN resistance to determine if the sequences in this *Rhg1* group can persist in greater than 3 copies. This information, coupled with the relationship between larger numbers of *Rhg1* repeats and increased resistance, suggests a new strategy to improve SCN resistance through addition of *Rhg1* copies

There remains a need for improved assays that can inexpensively but accurately determine the copy number of *Rhg1* or other high copy-number loci that confer adaptive traits (Curtis et al., 2012; Maron et al., 2013; Stebbing et al., 2013). We initially utilized qPCR with genomic DNA templates for this purpose, but found it challenging to obtain precise results for copy numbers above approximately four. Fiber-FISH provided definitive data and whole genome sequencing provided accurate estimates of higher copy-number regions as long as the genome-wide read depth exceeded approximately a two-fold coverage. Comparative genome hybridization (CGH) methods can also be used (Roberts et al., 2012). However, these relatively complex procedures are not likely to be useful, for example, in a plant breeding germplasm screen that seeks to identify rare individuals or infrequent recombination events carrying usefully elevated copy numbers.

Biochemical characterization of the “wild-type” (Williams 82-type), low-copy and high-copy versions of the *Glyma18g02590*  $\alpha$ -SNAP alleles also is needed, to determine what if any altered functions they have compared to each other and to canonical  $\alpha$ -SNAP functions. We speculate that while the genomes containing the PI 88788-type  $\alpha$ -SNAP have apparently benefited from an increase in *Rhg1* copy number, the genomes with the Peking-type  $\alpha$ -SNAP may have remained at three copies because of selection against an unknown negative impact of the Peking-type full-length  $\alpha$ -SNAP. *Rhg1* copy number in these genomes may also be affected by the shorter splice isoform of the *Glyma18g02590*  $\alpha$ -SNAP that was only detected in the low-copy *Rhg1* lines. Alternatively, the loss of a wild-type (Williams 82-like)  $\alpha$ -SNAP coding sequence in the three-copy genomes may have limited expansion of the locus. It is also possible that interactions with a specific *Rhg4* allele may favor the *Rhg1* locus configurations found in the low-copy *Rhg1* haplotypes.

The identification of the different copy numbers at *Rhg1* also suggests a hypothesis regarding the relatively ineffectual nature of low-copy *Rhg1* in the absence of the resistance-conferring *Rhg4* allele (Brucker et al., 2005; Liu et al., 2012). In the absence of Peking-type *Rhg4*, the 3 copies of *Rhg1* now known to be present in low-copy lines such as Peking have been shown to be more resistant to SCN infection than single copy *Rhg1* lines, suggesting that this *Rhg1* can function independently of resistance-associated *Rhg4* alleles (Brucker et al., 2005; Liu et al., 2012). This raises the possibility that *Rhg4* combined with the high-copy *Rhg1* may provide a broader-spectrum SCN resistance, while the Peking-type *Rhg1* resistance could possibly be improved by increasing the copy number or expression level. Also, stacked deployment of both types of *Rhg1* in single soybean lines could attenuate the development of virulent nematode populations. This type of research is increasingly important given the slow but ongoing erosion of the widely deployed PI 88788-derived resistance (Niblack et al., 2008; Tylka et al., 2012).

Our data help to explain the overlaps observed by many SCN-resistance specialists when comparing different soybean accessions with regard to their spectrum of resistance to a range of different SCN populations. For example, the resistance spectra of the Hg Type Test lines PI 88788, PI 209332 and Cloud (PI 548316) correlate highly, as do those of Peking (PI 548402), PI 90763, PI 89772 and PI 438489B (Colgrove and Niblack, 2008). Those two groupings match the *Rhg1* DNA sequence, copy number and  $\alpha$ -SNAP groups discovered in the present study.

PI 437654 is recognized for its particularly high levels of resistance against diverse nematode populations (e.g., Colgrove and Niblack, 2008). However, we discovered near identity of PI 437654 *Rhg1* copy number and sequence to other, less broadly resistant *Rhg1* low-copy soybean lines. Although *Rhg1* makes one of the strongest contributions to PI 437654-derived

resistance (Webb, 2012), the present finding re-emphasizes the importance of identifying and cloning additional SCN resistance QTL from PI 437654 (Wu et al., 2009).

Current models for evolution by gene duplication are often applied to single gene duplicates. A fascinating and unusual element of *Rhg1* is that gene copy number selection occurred, and research hypotheses are being tested, for a ~30 kb block of four genes that encode completely dissimilar proteins, three of which have been shown to contribute (Cook et al., 2012) to the phenotype that apparently has driven selection. Determining the exact course of evolution of the *Rhg1* locus is difficult, but our data strongly suggest that the repeats in the low-copy and high-copy class have a common origin. It is not clear if the common *Rhg1*-resistant progenitor diverged from susceptible lines prior to duplication, or if the divergence occurred after duplication. Either scenario could account for the highly similar sequence and the identical repeat junction found between low- and high-copy *Rhg1* lines if repeat homogenization or gene conversion has played a role in the evolution of the *Rhg1* locus and caused the high sequence identity between repeats within single plant lines.

Our data suggest that multiple evolutionary forces could have differentially affected the different genes in the repeat. Two of the proteins encoded at *Rhg1* (*Glyma18g02580* and *Glyma18g02610*) have identical derived amino acid sequences within the repeats and between the resistant lines, which matches predictions for gene duplicates fixed by positive selection for increased dosage and having a low rate of non-synonymous to synonymous substitutions ( $K_N/K_S < 1$ ) (Innan and Kondrashov, 2010). However, the presence of non-synonymous substitutions in *Glyma18g02590* in both the low- and high-copy *Rhg1* lines, caused by different nucleotide polymorphisms, suggests a different evolutionary course, the duplication and divergence scenario that is applicable to many gene duplicates (Ohno, 1970). It is also interesting to note the

identification of a premature stop codon in one copy of *Glyma18g02600* in Peking, despite the highly similar SCN resistance between Peking and the other resistant lines in the low-copy class. This provides further evidence that *Glyma18g02600* is not required for full *Rhg1*-mediated resistance, and could be the first glimpse of pseudogenation (Lynch and Conery, 2000). Hence the different genes in the *Rhg1* repeat apparently represent different evolutionary trajectories.

The identification of *Rhg1* DNA regions that exhibit differential methylation between SCN-resistant and SCN-susceptible accessions adds an additional layer of complexity to control of phenotype expression at *Rhg1*, and probably to *Rhg1* locus evolution. 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 a mechanism to maintain normal physiology following gene duplication (Qian et al., 2010). In recent work on mammalian gene duplicates, increased DNA methylation of promoter regions has been significantly correlated with gene duplicates and silencing, suggesting a potential mechanism for the restoration of dosage imbalance (Chang and Liao, 2012). This mechanism has also been suggested to follow whole genome duplications, for example in soybean, where for a number of gene pairs, one copy of the paralogous pair was often found to have increased repressive methylation and decreased expression (Schmitz et al., 2013). Our observations for *Rhg1* may seem to be the opposite of this, because in SCN-resistant lines with multiple *Rhg1* copies, hypermethylation is observed at genes that exhibit increased transcript abundance. However, expression of the multi-copy *Rhg1* genes might be even greater in the SCN-resistant genomes if there were not methylation. Although beyond the scope of the present study, recent identifications of dynamic methylation changes in *Arabidopsis* following biotic stress (Downen et al., 2012; Yu et al., 2013) suggest the hypothesis that the differentially methylated cytosine regions found upstream of *Glyma18g02580*,

*Glyma18g02590*, and *Glyma18g02610* could result in lower constitutive expression and increased expression of these genes following nematode infection. Future experiments to test this hypothesis may reveal further mechanisms that provide increased fitness and thereby impact the evolution of gene copy number variation.

### 3.5 Materials & Methods

#### Estimating copy number and transcript abundance

To estimate the number of *Rhg1* copies present in the Hg Type Test 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 frozen 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).

The soybean lines previously defined to make up the Hg Type Test nematode test were chosen for analysis (Niblack et al., 2002). The lines are: PI 548402 (Peking), PI 88788, PI 90763, PI 437654, PI 209332, PI 89772, PI 548316 (Cloud). The other line used and referenced in this work is Fayette, which was developed by crossing Williams(2) with PI 88788. Progeny from this cross were backcrossed with Williams(2) while selecting for SCN resistance.

### **Transcript analysis**

To confirm the annotation of transcripts at *Rhg1*, rapid amplification of cDNA ends (RACE) PCR was performed for 3' analysis of *Glyma18g02590* (Supplemental Table IV 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.

### **fiber-FISH**

Fiber-FISH experiments were carried out using the same methods and probes as previously detailed (Cook et al., 2012), and *Rhg1* repeat copy number findings are based on the maximum number of copies observed in at least ten separate probe-hybridizing DNA fibers for a given plant genotype.

### **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 following Illumina guidelines. Paired end libraries for samples PI 209332, PI89772, and Cloud were sequenced on a single Illumina HiSeq 2000 lane producing reads of 101 bp sequenced from both ends of the fragment. 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 300 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.

## **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 reference genome (version 1.1, assembly 1.89) corresponding to the *Rhg1* interval on chromosome 18 (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 sequencing reads were aligned to the full Williams 82 reference genome (build 1.89; <http://www.phytozome.net/cgi-bin/gbrowse/soybean/>) 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 Test data processing, PCR duplicates were marked at the lane level prior to merging the sequencing runs (McKenna et al., 2010).

### **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 remove false variants, a filter was applied to remove variants not sequenced at least three times and having a quality score greater than 50. Variant files were annotated with the program SnpEff as documented (Cingolani et al., 2012).

### **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.

### **Network Analysis**

To determine *Rhg1* sequence relationships between soybean lines, we performed multiple sequence alignment using ClustalW2. The open reading frames for the genes *Glyma18g02580*, *Glyma18g02590*, *Glyma18g02600*, and *Glyma18g02610* including 200 bp of upstream promoter sequence were concatenated and aligned. The alignment was used in SplitsTree (version 4.13.1)

to construct a sequence network (Huson and Bryant, 2006). The analysis pipeline included Uncorrect P for distances and NeighborNet for network construction. Parsimony-Uninformative sites were excluded from the network.

### **Analysis of nematode resistance**

To determine the relationship between nematode resistance and lines containing different copy numbers of *Rhg1*, we analyzed data collected as part of the Northern Regional Soybean Cyst Nematode Test (Cary and Diers, 2010; Cary and Diers, 2011, 2012; Cary and Diers, 2013). In total, we analyzed data from greenhouse nematode trials conducted on the 7 Hg-type soybean lines and the susceptible control line Lee for 78 SCN field populations. Six plants per genotype were tested against the 78 different nematode populations. To more accurately estimate the variance for Female Index, we performed random replacement using the software R (Team, 2009) with 1000 bootstrap replicates per genotype-nematode combination to estimate the variance. An ANOVA was computed using a linear mixed effect model (lmer) with bootstrap variances used to weight observations, expressed as the inverse of the variance. Residuals were checked for normality. P-values were calculated using the generated T-values, and a Bonferroni correction was applied to account for false positives resulting from multiple testing.

### **Methylation Analysis**

#### *Restriction Enzyme-Based Methylation Discovery*

Locus specific DNA methylation was analyzed using the methylation specific endonuclease McrBC. McrBC digests DNA with methylated cytosines in a sequence independent manner while unmethylated DNA is unaffected. 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 did not produce the intended product following PCR because the template DNA was digested by McrBC. DNA that was not methylated or not treated with the enzyme yielded a product of the expected size.

#### *Computational Methylation Analysis*

Data were downloaded from National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) Series accession number GSE41753, previously deposited (Schmitz et al., 2013). These data were analyzed using custom scripts written in Java or Bash to compute the data, and results are presented in Figure 8, and Supplemental Figures S3, S4 and S5.

To estimate *Rhg1* copy number, sequence from GEO accession number GSE41753 (GEO number GMS1024005 through GMS1024008, GMS1134684, GMS1134698 through GMS1134700, GMS1134705, GMS1134706, GMS1134709, GMS1134712 through GMS1134714, GMS1134716, GMS1134718, GMS1134720, GMS1134722, GMS1134723, GMS1134729 through GMS1134732, GMS1134734, GMS1134736, GMS1134741, GMS1134744, GMS1134749 and GMS1134756) were analyzed. The total number of cytosine sequencing reads were 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, Supplemental Figure S3A.

To determine single base cytosine methylation at the *Rhg1* locus, sequences from GEO accession number GSE41753 were used for the corresponding groups: Parental Lines (GEO# GSM1024005 and GSM1024006); Single-copy *Rhg1* progeny (GEO# GSM1024007, GSM1134698 through GSM1134700, GSM1134709, GSM1134712, GSM1134714, GSM1134716, GSM1134720, GSM1134723, GSM1134729 through GSM1134731, GSM1134734, GSM1134741, and GSM1134749); Three-copy *Rhg1* progeny (GEO# GSM1024008, GSM1134684, GSM1134713, GSM1134732, GSM1134744, and GSM1134756). The total number of cytosine sequencing reads and the total number of cytosine sequencing reads supporting methylation were summed over 150bp bins starting at position 1,626,000 through 1,668,000 for a total of 280 bins. For each bin, the methylation level was computed by dividing the total number of cytosine reads supporting methylation by the total number of cytosines sequenced. Methylation levels were computed in the CG, CHG, and CHH sequence context. The data are represented in Figure 8 and Supplemental Figures S4 and S5.

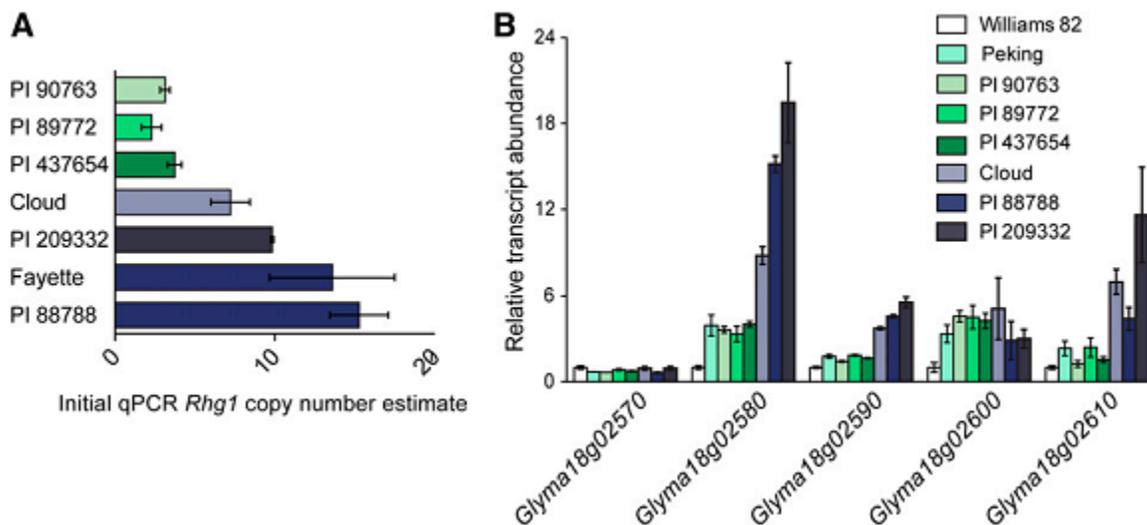
To estimate expression of genes within and adjacent to the *Rhg1* repeat, processed RNA-sequencing data were used to compare transcript levels across the 4 tested genotypes (GEO series GSE41753\_RPKM supplementary file). To assess transcription differences, the reads per kilobase per million mapped sequence reads (RPKM) values from the 3 replicates of the single-copy *Rhg1* 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.



### **3.6 Acknowledgments**

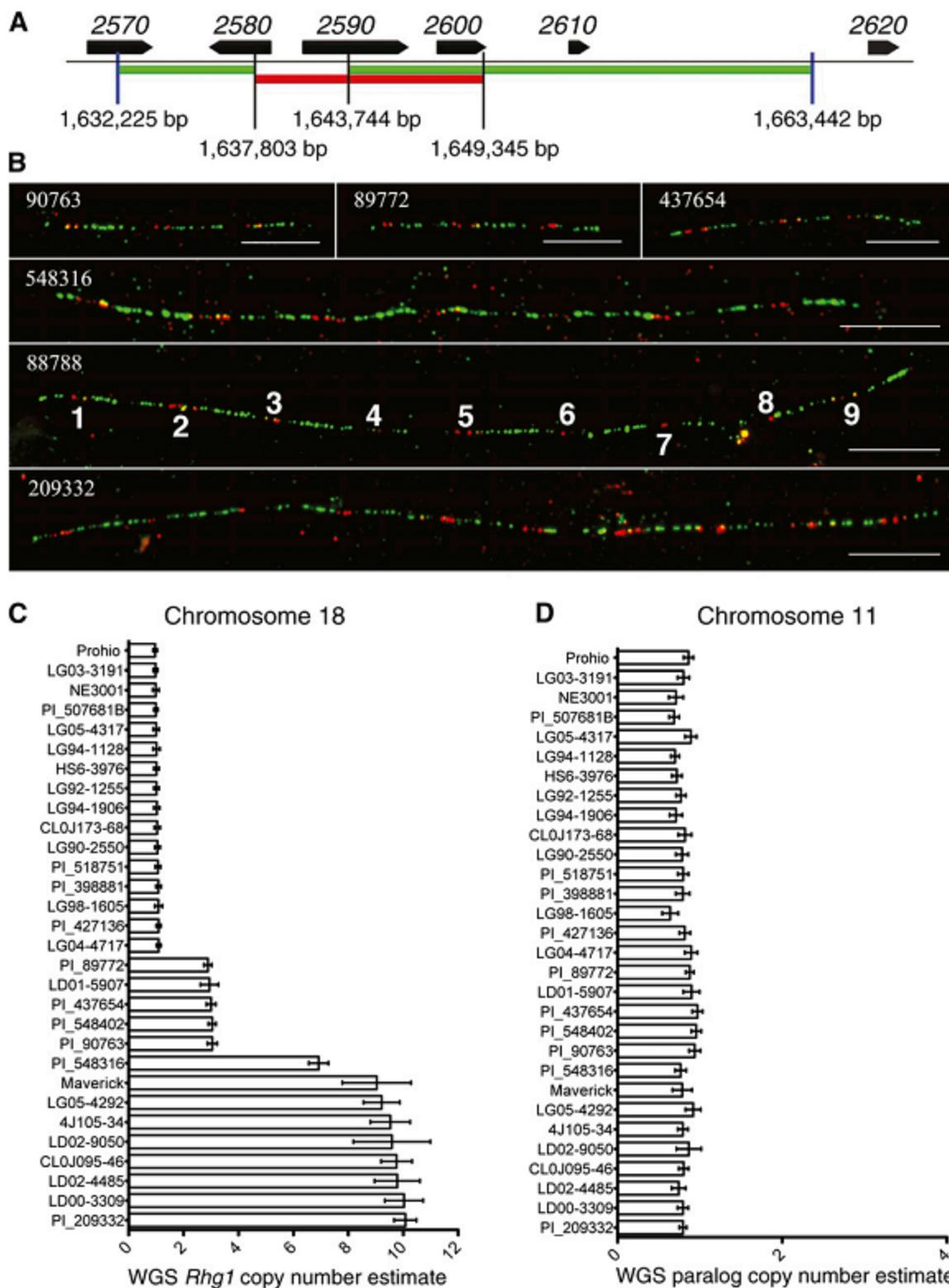
We thank Matthew Hudson and Tong Geon Lee for multiple conversations about this work and sharing of results prior to publications, and Guy Plunkett, Marie Adams, John Alliet, Xiao-yu Liu and the University of Wisconsin-Madison Biotechnology DNA Sequencing and Bioinformatics Resource Center for education and support.

### 3.7 Figures



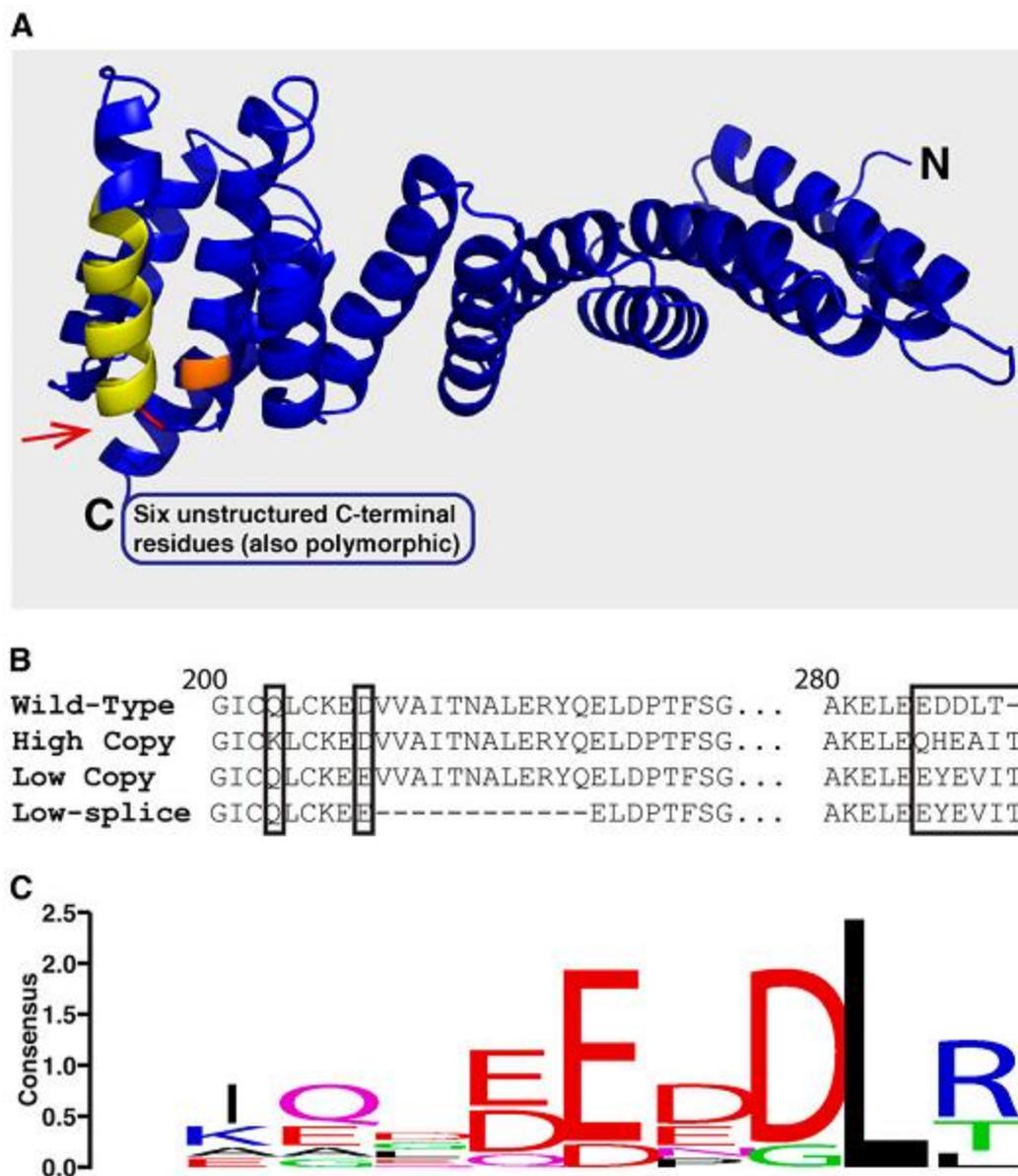
**Figure 1. Estimates of copy number and transcript abundance suggest different types of *Rhg1* loci.**

(A) Initial *Rhg1* copy number estimates, obtained using qPCR to amplify genomic DNA, identify two groups of SCN-resistant lines: low-copy number, between 2 and 4 repeats (PI 90763, PI 89772, PI437654) and high-copy number estimated to have greater than 7 repeats (Cloud, PI209332, Fayette, PI 88788). Copy number is expressed as ratio of qPCR template abundance estimates for *Rhg1* repeat junction and for a non-duplicated neighboring gene. (B) Transcript abundance, relative to SCN-susceptible Williams 82, indicates the presence of two expression groups of *Rhg1* loci in SCN-resistant lines. Roots from lines identified in subsequent work as having 3 copies of the 31 kb *Rhg1* repeat (Peking, PI90763, PI89772, PI437654) exhibit lower transcript abundance than lines with 7, 9 or 10 *Rhg1* copies (Cloud, PI88788, PI209332). The one complete copy of *Glyma18g02570*, located immediately outside of the *Rhg1* repeat region (Cook *et al.* 2012 and present work), is expressed at a similar level across all the tested lines. Expression level of *Glyma18g02600* is near the detection limit for qPCR.



**Figure 2. Whole genome re-sequencing and fiber-FISH define the copy number of *Rhg1* in Hg type lines.**

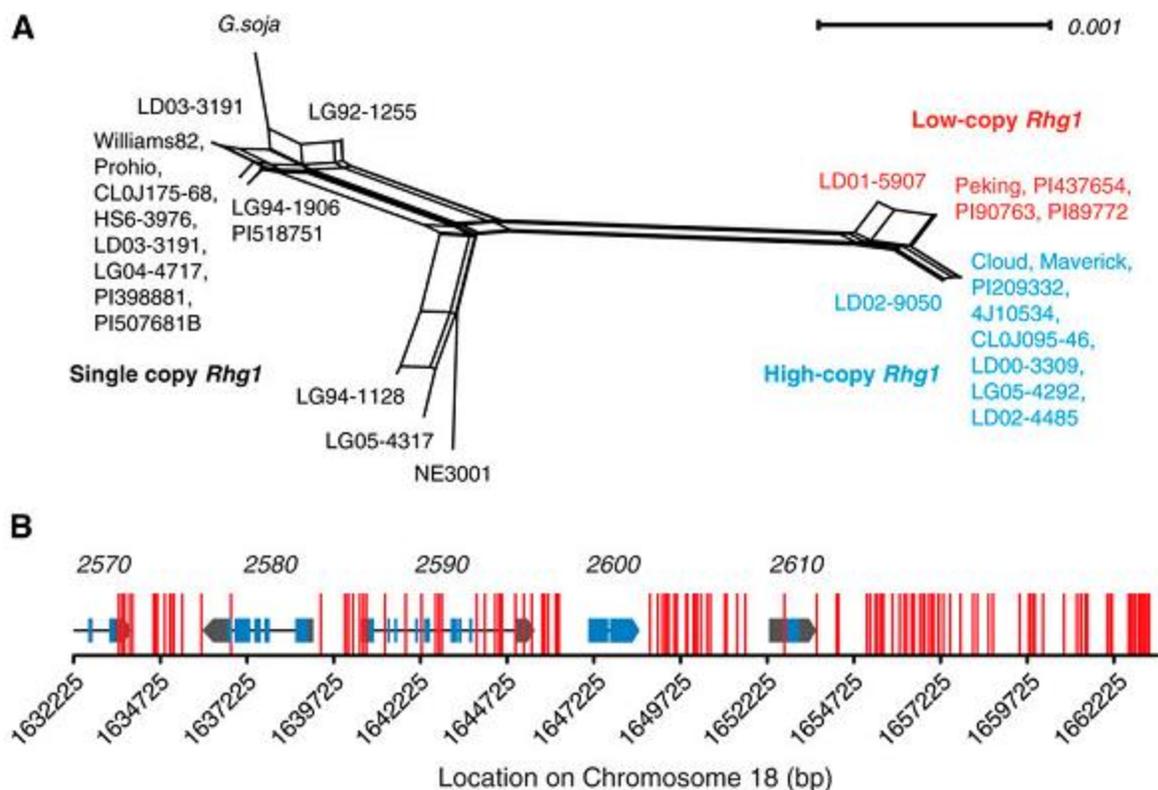
(A) Diagram of red (11.5 kb) and green (25.3 kb) DNA probes used to detect *Rhg1* repeats in fiber-FISH. Gene and bp numbers are from chromosome 18 of the soybean Williams 82 reference genome. (B) Representative fiber-FISH images collected from six Hg Type Test soybean lines. As previously documented for three soybean lines (Cook et al. 2012), the *Rhg1* locus is present as multiple direct repeats on single DNA fibers. The present data indicate a copy number of 3 for PI 90763, PI 89772 and PI 437654, and copy numbers of 7, 9 and 10 for PI 548316 (Cloud), PI 88788 and PI 209332, respectively. The repeats are labeled for clarity for the representative fiber shown in box 88788 (PI 88788). White bar = 10  $\mu$ m in each panel. (C) *Rhg1* copy number for 30 soybean lines, based on whole genome sequence read depth analysis. Average read depth was determined for 1 kb bins across the *Rhg1* repeat and for 30 kb on each side of the *Rhg1* repeat region. Data for the flanking single-copy regions from a given line were used to normalize the read depth data of 1kb bins within the *Rhg1* repeat to determine copy number (mean  $\pm$  standard error of the mean). (D) Copy numbers determined as in panel (C), but for the *Rhg1* paralog locus on chromosome 11.



**Figure 3. Resistant type *Rhg1* classes encode unique  $\alpha$ -SNAP alleles with polymorphisms in highly conserved residues localized at the C-terminus**

(A) Structure of Glyma18g02590 from Williams82 modeled to the crystal structure of yeast  $\alpha$ -SNAP (sec17p, PDB 1QQE). The Q203K substitution unique to high-copy *Rhg1* encoded  $\alpha$ -SNAPs is colored orange. The D208E substitution present only in low-copy *Rhg1*  $\alpha$ -SNAPs is shown in red (red arrow points to this residue). An alternative splice isoform detected in low-

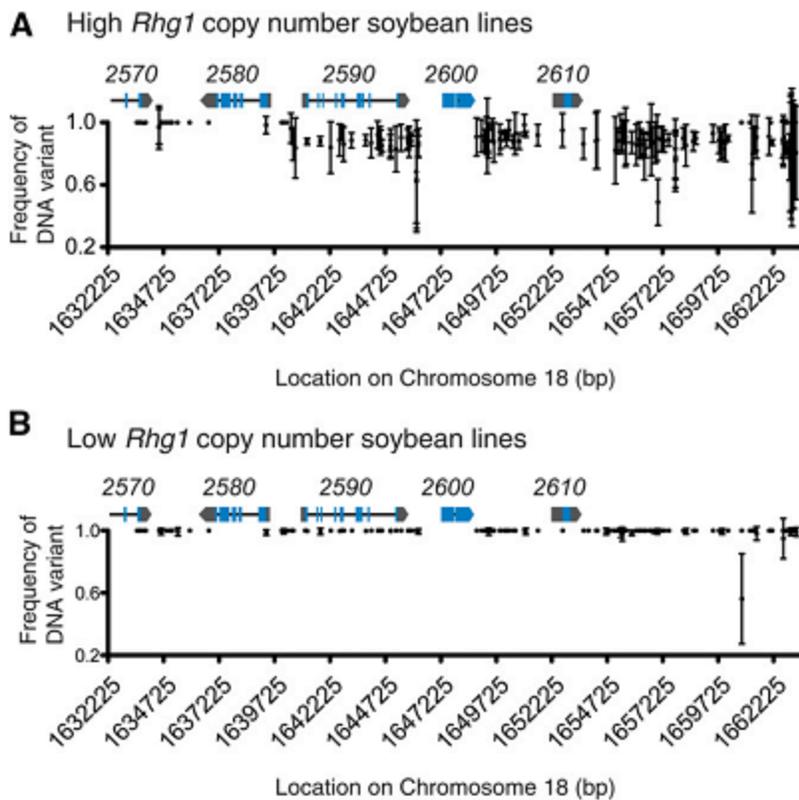
copy *Rhg1* classes removes 12 residues from the full-length protein (displayed in yellow). In both *Rhg1* resistant-class  $\alpha$ -SNAPs, similar but distinct polymorphisms in the final six C-terminal residues are present (these polymorphic residues not modeled; final residues predicted to be unstructured in PDB 1QQE). (B) Amino acid sequence of Glyma18g02590 from Williams82 (susceptible) aligned to predicted amino acid sequences of both high and low copy class *Rhg1*  $\alpha$ -SNAPs. Note that the low-copy splice isoform is predicted to exclude residues 209 – 220. No predicted amino acid polymorphisms in Glyma18g02590 from the sequenced SCN susceptible lines have been detected. (C) Logo displaying the consensus sequence for the final ten C-terminal residues of  $\alpha$ -SNAP from eight diverse eukaryotes (*H. sapiens*, *D. melanogaster*, *S. cerevisiae*, *C. elegans*, *D. rerio*, *B. taurus*, *A. thaliana*, *G. max*). Strikingly, the unique C-terminal polymorphisms discovered in *Rhg1* resistant type  $\alpha$ -SNAPs occur at these five most highly conserved residues.



**Figure 4. Network analysis and shared polymorphisms support three *Rhg1* locus types and a single SCN-resistant-type progenitor**

(A) Neighbor-net analysis indicates two distinct groups based on *Rhg1* sequence, separating the SCN susceptible lines (left, single-copy *Rhg1*) and the SCN resistance lines (right, multi-copy *Rhg1*). The resistant lines further split into two groups that correspond with *Rhg1* copy number (see Figs. 1 and 2); lines containing 3 *Rhg1* copies are noted in red and those containing  $\geq 7$  copies are noted in blue. The four coding genes within the *Rhg1* repeat, including 200 bp of sequence upstream of the start codon were used for analysis. (B) DNA variant sites present in all seven Hg Type Test soybean lines (low-copy and high-copy number *Rhg1*; SCN-resistant) but absent from all sequenced SCN-susceptible single-copy *Rhg1* lines. Vertical red lines show locations of these 148 DNA variant sites, which are 75% of all of the 197 SNP or INDEL DNA variant sites present in at least one Hg Type Test line but not present in any of the examined

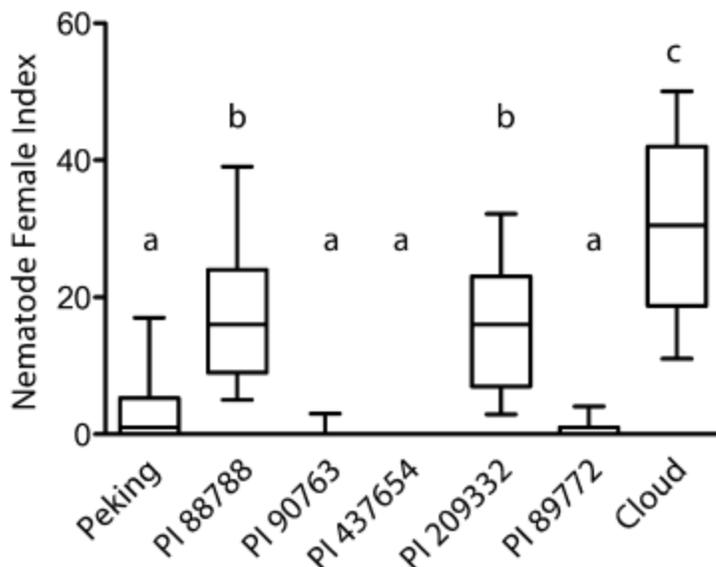
SCN-susceptible lines. *Rhg1* locus gene models shown at correct x-axis position for reference (blue exons, black line introns, grey untranslated regions); gene name is above gene model (e.g., 2570 = *Glyma18g02570*).



**Figure 5. The frequency of DNA variant sites across *Rhg1* repeats reveals heterogeneity between repeats in high-copy but not low-copy *Rhg1* containing lines.**

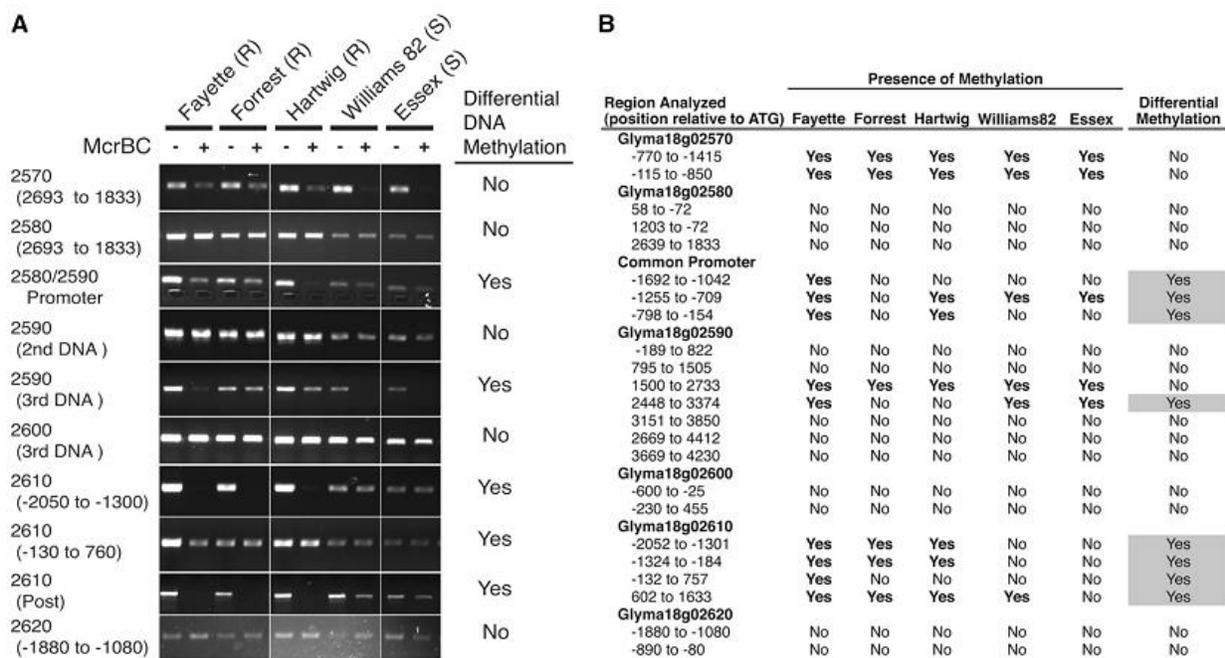
(A) Nearly homogeneous presence of same non-Williams 82 DNA sequence for variant sites in all copies (left quarter) or all but one copy (right three-quarters) of the *Rhg1* repeat. X-axis: Location of DNA variant site (SNP or INDEL) within *Rhg1* locus on soybean chromosome 18. Y-axis: Proportion of all DNA sequence reads with variant (high-copy-type) sequence rather than the reference Williams 82-type (single-copy *Rhg1*) sequence, at the designated DNA variant site. Data combined for the three *Rhg1* high-copy class Hg Type Test soybean lines LD00-3309 (PI 88788), PI 209332 and Cloud; mean frequency and std. err. of mean for the three soybean lines are shown. *Rhg1* locus gene models shown at correct x-axis position for reference (blue exons, black line introns, grey untranslated regions); gene name is above gene model (e.g., 2570 = *Glyma18g02570*). (B) Near identity of the three repeats in *Rhg1* low-copy lines, and absence

of a Williams 82-type segment. Figure as in (A) except showing combined data for the four *Rhg1* low-copy class Hg Type Test soybean lines Peking, PI 90763, PI 437654 and PI 89772.



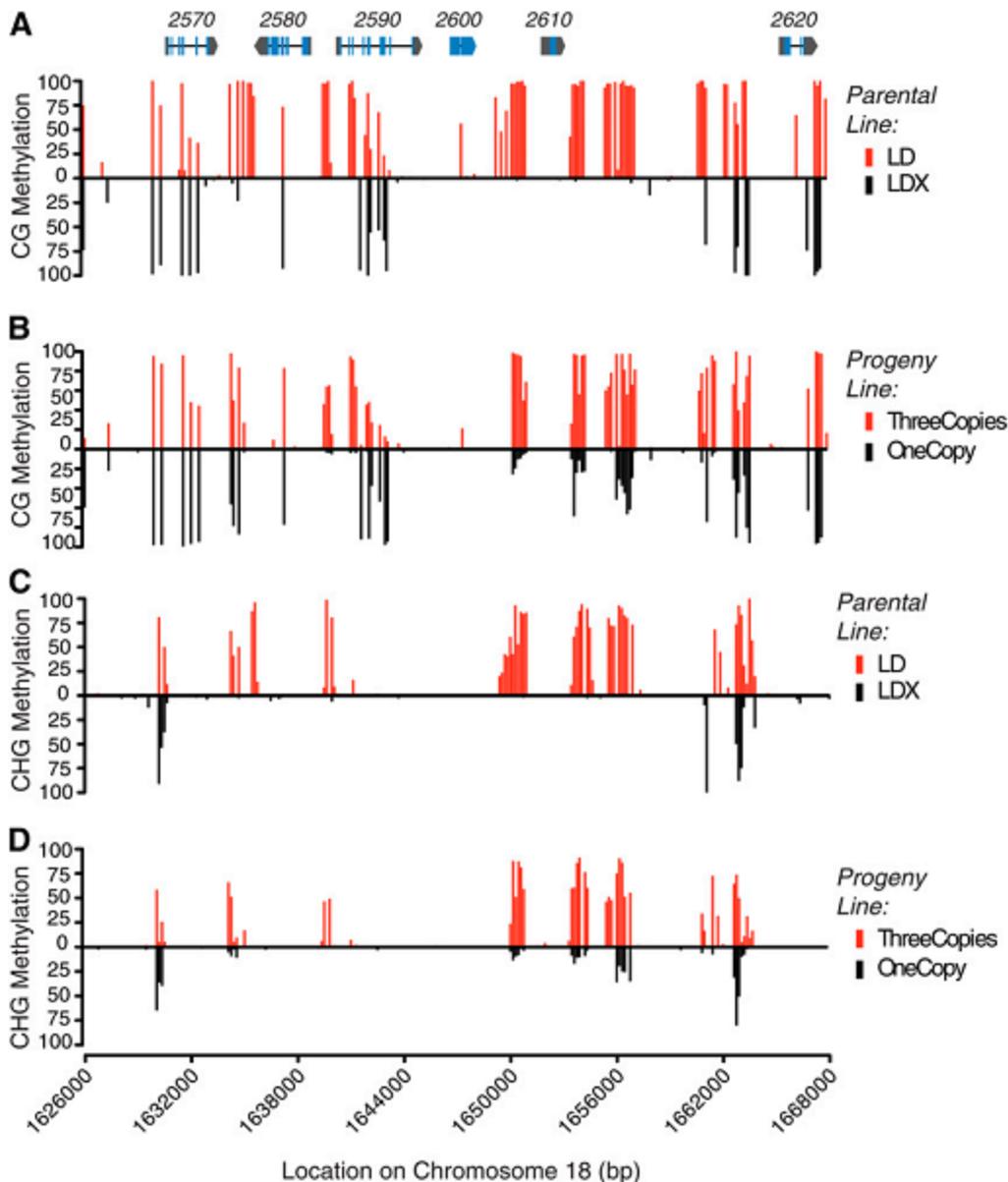
**Figure 6. Nematode resistance data from 78 diverse SCN populations indicates similarities in resistance profiles based on copy number.**

Nematode development data obtained for the seven Hg Type Test SCN-resistant lines, for greenhouse assays conducted as part of the 2009-2012 Northern Regional SCN Tests. Data analyzed for 78 *H. glycines* nematode populations collected from 12 U.S. states and Canada provinces. Female index is the percentage of SCN cysts that developed on the resistant soybean line relative to the susceptible control soybean line. Boxes show median and 25%-75% range of data; whiskers extend to 10% and 90% of the data. For statistical analysis, variance was calculated by random replacement with 1000 bootstrap replicates for each line within a given nematode population (see methods). This calculated variance was used in a weighted ANOVA; soybean lines not sharing the same letter above whisker had significantly different means following Bonferroni correction for multiple testing, at a p-value < 0.001.



**Figure 7. Differential *Rhg1* locus DNA methylation between SCN-resistant and SCN-susceptible lines, particularly in control regions upstream of SCN resistance genes.**

(A) Representative gel images of PCR products from soybean root genomic DNA template treated with restriction endonuclease McrBC (+) or buffer only (-) prior to PCR. McrBC cleaves (G/A)<sup>m</sup>C sites containing methylcytosine, preventing PCR amplification of cleaved template strands so that PCR product abundance goes down with increasing levels of methylation. “Differential DNA Methylation” scored as positive if any soybean line differed from other lines in McrBC-sensitivity of the PCR product in two independent tests. Soybean lines are denoted as either resistant (R) or susceptible (s) to SCN. (B) Summary table for replicated McrBC study described in (A) with 23 PCR primer pairs used to assess DNA methylation within the *Rhg1* locus. The presence of methylation is listed as yes if both DNA samples showed reduced PCR amplification following McrBC DNA treatment. Right column reports methylation differences between different soybean lines.



**Figure 8. DNA methylome sequence from three-copy and single-copy *Rhg1* lines and their progeny further define differential methylation at *Rhg1* SCN resistance genes.**

Levels of DNA methylation reported as proportion of methylated cytosines detected from bisulfite sequencing. Data are for 150bp bins represented by a single vertical line. *Rhg1* locus gene models are shown at the top of panel (A) at correct x-axis position along chromosome 18 shown below panel (D) for reference (blue exons, black line introns, grey untranslated regions);

gene name above gene model (e.g., 2570 = *Glyma18g02570*). (A) Levels of cytosine methylation for the sequence context CG, showing differential methylation of parental line LD (three-copies of *Rhg1*, red vertical lines above x-axis) relative to parental line LDX (single copy of *Rhg1*, black vertical lines below x-axis). The greatest differential methylation is present up and downstream of the *Glyma18g02580* open read frame (ORF), in the common promoter for *Glyma18g02580* and *Glyma18g02590*, and both up and downstream of the *Glyma18g02610* ORF, with more methylation in the three-copy *Rhg1* SCN-resistant line. (B) Average CG methylation in F3-derived progeny families of the cross between lines LD and LDX, either for all six progeny estimated to have an *Rhg1* copy number of 3 (red vertical lines above x-axis), or for all 16 progeny lines estimated to have an *Rhg1* copy number of 1 (black vertical lines below x-axis). Substantial similarities to the parental CG methylation patterns are evident. (C), (D) Levels of cytosine methylation for the sequence context CHG, where H can be either an A,T, or C. (C) Analysis similar to (A) except for CHG sequence context. The same regions identified as differentially methylated in (A) are again identified as hypermethylated. (D) Analysis similar to (B) except for CHG sequence context, using the same progeny as (B).

## 3.8 Tables

**Table 1.** Summary statistics for DNA variant analysis at *Rhg1* from whole-genome sequencing shows higher rates of polymorphism in SCN-resistant lines

Copy No., *Rhg1* copy number estimated from sequencing read depth or from fiber-FISH (asterisks). Variant Class, Whole-genome sequencing for 30 soybean lines and one *G. soja* line was analyzed for DNA variants and classified as SNPs, insertion, or deletion relative to the Williams 82 reference genome. The total number of DNA variants of each type across the 31-kb *Rhg1* sequence are reported. Total, Sum of the SNP, insertion, and deletion variants. Variant Location columns report numbers of variants in each type of genome region. UTR, Untranscribed region.

Genotype	Variant Class				Variant Location				
	Copy No.	SNPs	Insertion	Deletion	Total	Exon	Intron	Intergenic	UTR
LD00-3309	10.03	190	36	32	258	6	24	220	8
LG05-4292	9.23	195	38	34	267	6	25	228	8
4J105-34	9.53	196	35	32	263	6	24	225	8
CL0J095-46	9.76	195	36	32	263	6	24	225	8
LD02-4485	9.79	189	36	33	258	6	25	219	8
LD02-9050	9.56	177	32	28	237	6	24	199	8
Maverick	9.04	187	33	31	251	6	24	213	8
Cloud	7*	194	35	32	261	6	25	222	8
PI 209332	10*	197	35	33	265	6	27	224	8
PI 437654	3*	193	37	35	265	5	27	224	9
Peking	3*	200	38	33	271	6	25	232	8
PI 89772	3*	194	37	36	267	5	25	229	8
PI 90763	3*	192	37	34	263	5	25	224	9
LD01-5907	2.95	146	18	17	181	5	24	146	6
NE3001	0.99	63	3	6	72	4	23	42	3
LG05-4317	1.01	72	4	5	81	1	19	60	1
LG94-1128	1.02	45	2	3	50	1	15	33	1
LG92-1255	1.02	46	6	3	55	0	3	51	1
PI 518751	1.07	44	2	3	49	0	0	49	0
LG94-1906	1.03	26	0	2	28	0	0	28	0
LG03-3191	0.98	28	0	0	28	0	0	28	0
CL0J173-68	1.04	0	0	0	0	0	0	0	0
HS6-3976	1.02	0	0	0	0	0	0	0	0
LG04-4717	1.10	0	0	0	0	0	0	0	0
PI 398881	1.09	0	0	0	0	0	0	0	0
PI 427136	1.09	0	0	0	0	0	0	0	0
PI 507681B	1.00	0	0	0	0	0	0	0	0
LG90-2550	1.05	0	0	0	0	0	0	0	0
Prohio	0.97	0	0	0	0	0	0	0	0
LG98-1605	1.09	0	0	0	0	0	0	0	0
<i>G. soja</i>	1.10	42	3	4	49	1	3	45	0

**Table 1.**

**Table II.** Amino acid polymorphisms for genes encoded within and adjacent to the *Rhg1* repeat, from all analyzed soybean lines

Position, Chromosome 18 base-pair position relative to the Williams 82 reference genome, with gene names (and putative gene product functions in parentheses) above the relevant base-pair positions. The remaining columns indicate soybean accessions. Low-Copy Lines, All five soybean lines estimated to contain three copies of *Rhg1* have the same amino acid polymorphism and are represented by a single column. High-Copy Lines, All lines estimated to contain seven or more copies of *Rhg1* have the same amino acid polymorphism and are represented by a single column. Only three additional soybean lines contain amino acid polymorphisms for any of the six genes analyzed and are listed in individual columns. Amino acid polymorphisms are reported as the amino acid present in Williams 82, the amino acid position, and the resulting new amino acid discovered. Not shown here is the soybean line Peking, which contains a single-nucleotide deletion in some *Rhg1* repeat copies that introduces a frame shift at amino acid 214 and results in a premature stop codon after amino acid 222 of *Glyma18g02600* (eliminating 58% of predicted protein).

Position	Low-Copy Lines	High-Copy Lines	NE3001	LG05-4317	LG94-1128
<i>Glyma18g02570</i> (unknown function)			No polymorphisms		
<i>Glyma18g02580</i> (amino acid transporter)					
1638989			V9I		
<i>Glyma18g02590</i> ( $\alpha$ -SNAP)					
1643208		Q203K			
1643225	D208E				
1644965		E285Q			
1644968	D286Y	D286H			
1644972	D287EV	D287EA			
1644974	L288I	L288I			
<i>Glyma18g02600</i> (PLAC-8 family)					
1647134			V23A		
1647764	214 stop (Peking)				
1648561				M480L	M480L
<i>Glyma18g02610</i> (unknown function)			No polymorphisms		
<i>Glyma18g02620</i> (SEL-1-like)			No polymorphisms		

**Table 2.**

## 3.9 Supplemental Figures

```

Ch18_Williams      1  GGTGGGGCTTGTGGCTCCAAGTATGAAGATGCCCGCATCTCTCGATAAAGCCGCC
Chr18_Peking      1  .....
Chr11_Williams    1  .....C.....T.....
Truncated_alphaSNAP 1  .....
Chr11_Peking      1  .....C.....

Ch18_Williams    61  AATGCTTCAAGCTCGCCAAATCATGGGACAAGGCTGGAGCGACATACCTGAAGTTGGCA
Chr18_Peking     61  .....
Chr11_Williams   61  .....A.....
Truncated_alphaSNAP 61  .....
Chr11_Peking     61  .....A.....

Ch18_Williams   121  AGTTGTCATTTGAAGTTGGAAAGCAAGCATGAAGCTGCACAGGCCCATGTCGATGCTGCA
Chr18_Peking    121  .....
Chr11_Williams  121  .....T.....
Truncated_alphaSNAP 121  .....
Chr11_Peking    121  .....T.....

Ch18_Williams   181  CATTGCTACAAAAGACTAATATAAACGAGTCTGTATCTTGCTTAGACCGAGCTGTAAT
Chr18_Peking    181  .....
Chr11_Williams  181  ..A...T...A.....A..C.....
Truncated_alphaSNAP 181  ..A.....A.....
Chr11_Peking    181  ..A...T...A.....A..C.....

Ch18_Williams   241  CTTTCTGTGACATTGGAAGACTCTCTATGGCTGCTAGATATTTAAAGGAAATGCTGAA
Chr18_Peking    241  .....
Chr11_Williams  241  .....A.....
Truncated_alphaSNAP 241  .....
Chr11_Peking    241  .....A.....

Ch18_Williams   301  TTGTACGAGGGTGAACAGAATATTGAGCAGGCTCTTGTTTACTATGAAAAATCAGCTGAT
Chr18_Peking    301  .....
Chr11_Williams  301  ....T.....
Truncated_alphaSNAP 301  .....
Chr11_Peking    301  ....T.....

Ch18_Williams   361  TTTTTCAAAATGAAGAAGTGACAACCTCTGCGAACCAATGCAAAACAAAAGTTGCCAG
Chr18_Peking    361  .....
Chr11_Williams  361  .....A.....
Truncated_alphaSNAP 361  .....
Chr11_Peking    361  .....A.....

Ch18_Williams   421  TTTGCTGCTCAGCTAGAACAATATCAGAAGTCGATTGACATTTATGAAGAGATAGCTCGC
Chr18_Peking    421  .....
Chr11_Williams  421  .....G...C.....
Truncated_alphaSNAP 421  .....G.....
Chr11_Peking    421  .....G...C.....A.....

Ch18_Williams   481  CAATCCCTCAACAATAATTTGCTGAAGTATGGAGTTAAAGGACACCTTCTTAATGCTGCC
Chr18_Peking    481  .....
Chr11_Williams  481  .....G.....
Truncated_alphaSNAP 481  .....
Chr11_Peking    481  .....G.....

Ch18_Williams   541  ATCTGCCAACTCTGTAAAGAGGACGTTGTTGCTATAACCAATGCATTAGAACGATATCAG
Chr18_Peking    541  .....G.....
Chr11_Williams  541  .....G...T...A...G.....
Truncated_alphaSNAP 541  .....G...A...G.....
Chr11_Peking    541  .....G...T...A...G.....

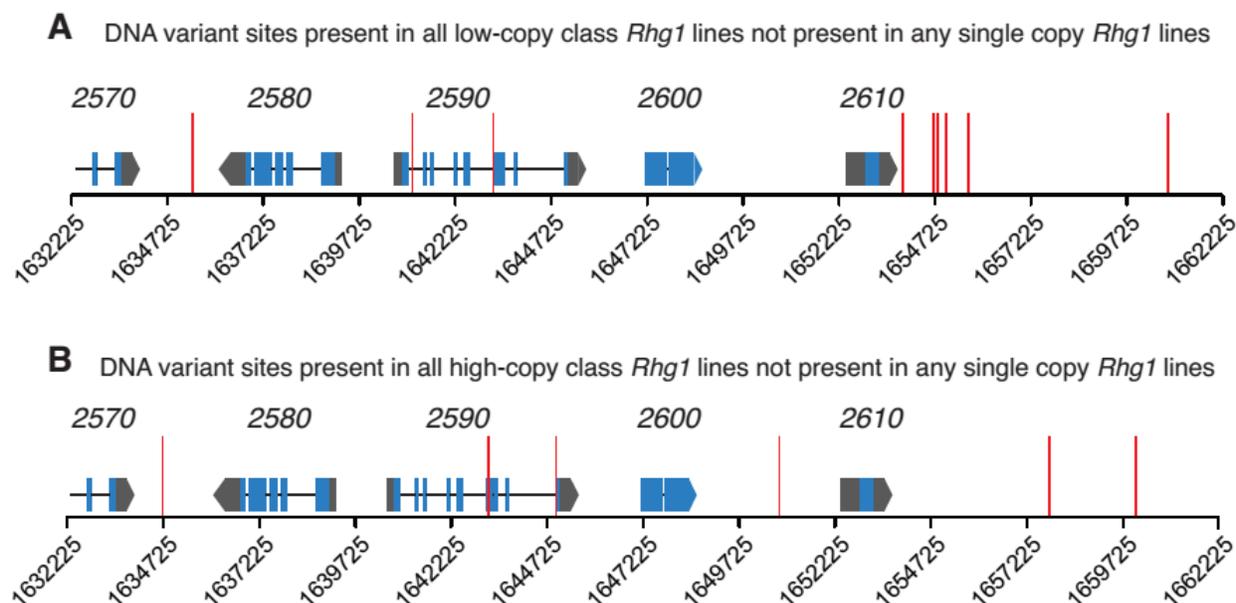
Ch18_Williams   601  GAACTGGATCCAACATTTTCAGGAACACGTGAATATAGATTGTTGGCGGACATTGCTGCT
Chr18_Peking    601  .....
Chr11_Williams  601  .....G.....T.....
Truncated_alphaSNAP 601  .....T.....TTAGG.CACTAG
Chr11_Peking    601  .....T.....TTAGG.CACTAG

Ch18_Williams   661  GC
Chr18_Peking    661  ..
Chr11_Williams  661  ..
Truncated_alphaSNAP --
Chr11_Peking    --

```

**Figure S1. Previously reported truncated allele of  $\alpha$ -SNAP shares higher sequence similarity to the paralog encoded on chromosome 11 and is likely not encoded by *Glyma18g02590* at *Rhg1*.**

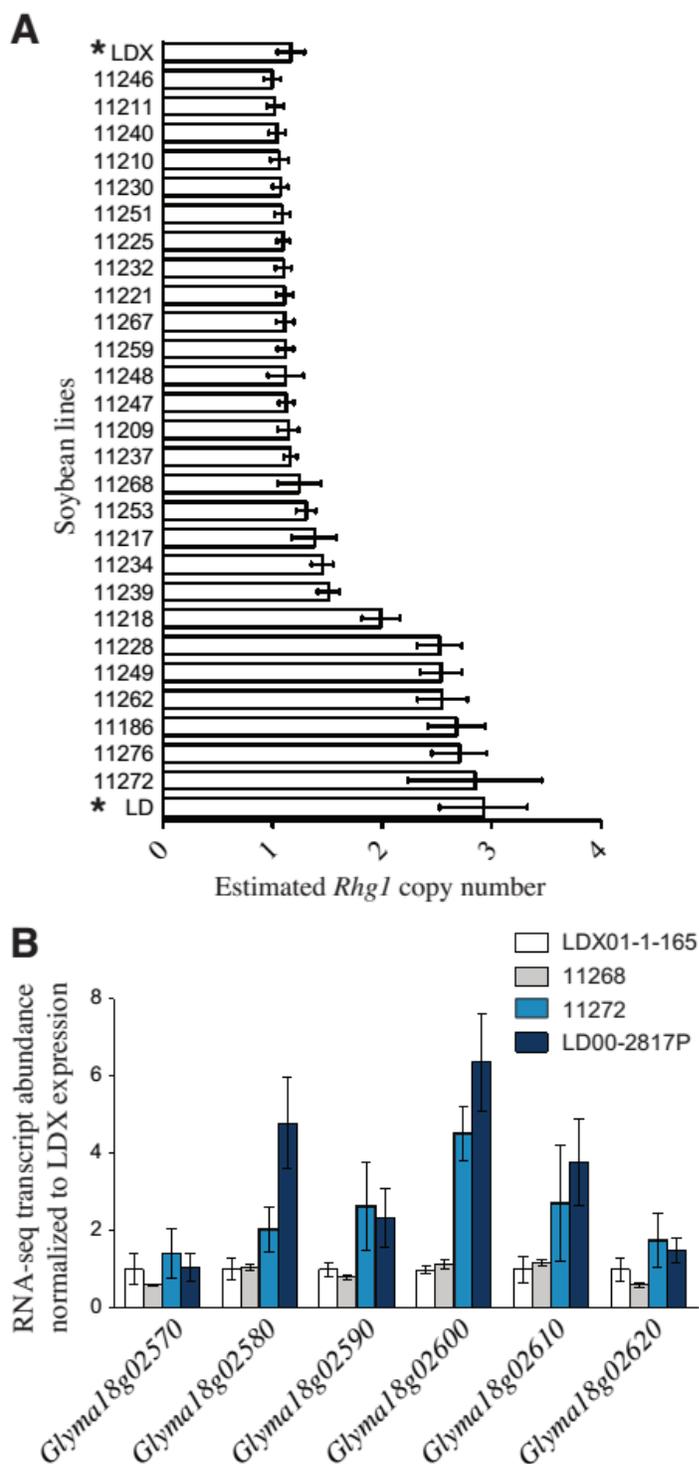
Nucleic acid alignment for the first 661 bases of  $\alpha$ -SNAP encoded by chromosome 18 (*Rhg1*) and 11 (paralog) from Williams82 and Peking, and the previously reported truncated allele sequence in (Matsye et al., 2012). Sequence from Williams82 is shown and positions with an identical sequence are listed as (.). The sequence reported in (Matsye et al., 2012) for the truncated allele of *Glyma18g02590* is most similar to the Williams 82 and Peking paralogs encoded on chromosome 11. The polymorphism reported to change the exon-intron boundary and cause the splice variant is highlighted in yellow, and the resulting in frame stop codon is highlighted in red.



**Figure S2. The DNA sequence of *Rhg1* repeats has continued to diverge between the low- and high-copy containing lines.**

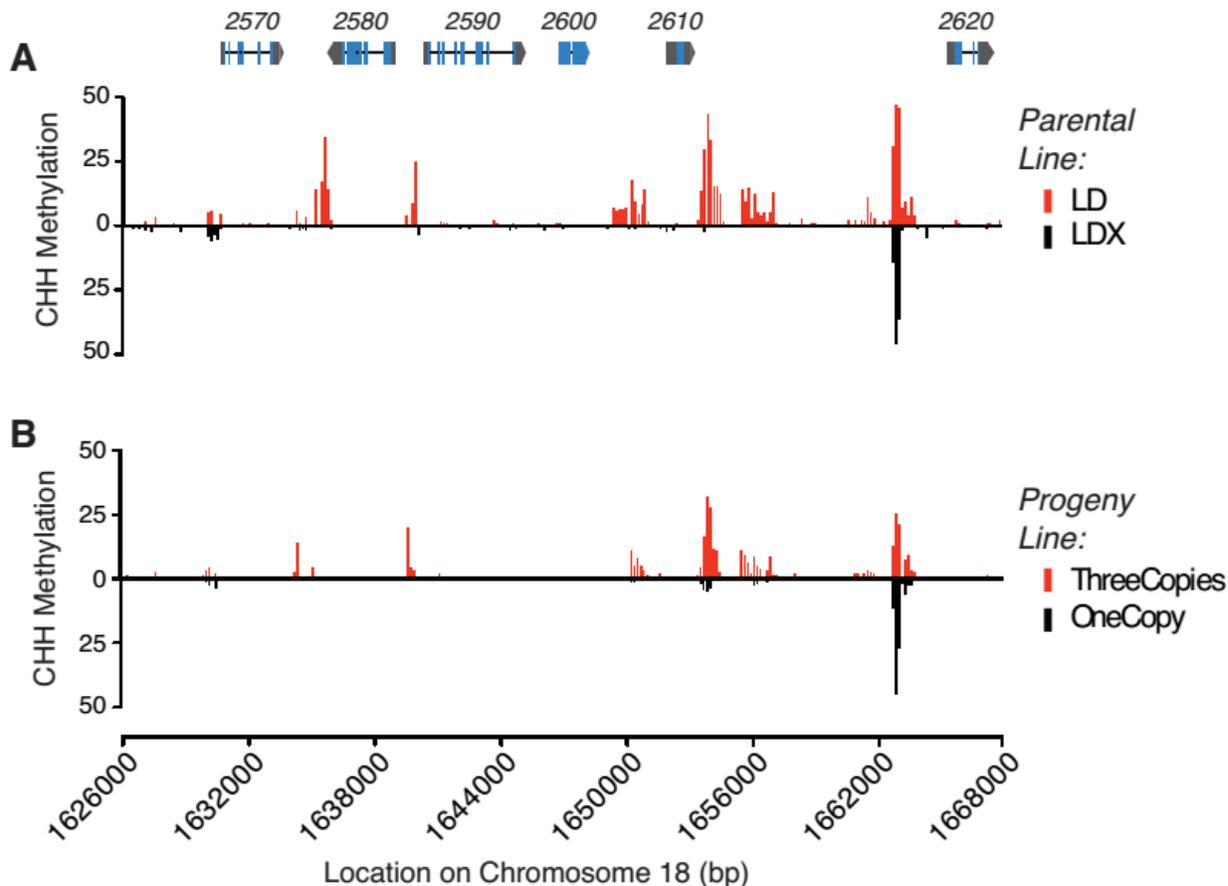
(A) DNA variant sites present in low-copy *Rhg1* Hg Type Test lines indicate that following divergence from the high-copy group, the locus is continuing to evolve. There are 10 DNA variants present in all low-copy *Rhg1* lines, not present in the high-copy or single-copy lines.

(B) DNA variant sites are shown as in (A), but instead only DNA variant sites present in high-copy Hg Type Test lines, not present in low-copy or single-copy lines. Red vertical bars represent the location of the DNA variants across the *Rhg1* locus. *Rhg1* locus gene models shown at correct x-axis position for reference (blue exons, black line introns, grey untranslated regions); gene name is above gene model (e.g., 2570 = *Glyma18g02570*).



**Figure S3. Copy number estimates and RNA-seq analysis indicate the presence of multiple copies of *Rhg1* in SCN resistant parent LD00-2871P and some progeny with a concomitant increase in transcription.**

(A) Parental lines and 27 progeny were analyzed for *Rhg1* copy number estimates based on cytosine sequencing depth. The parental line LD00-2871P (LD) is estimated to contain 3 copies of the *Rhg1* repeat, consistent with its derivation of SCN resistance from PI 437654 described here. The two parental lines are denoted with \*. (B) Relative transcript abundance based on RNA-seq reads indicates the 4 genes transcribed within the *Rhg1* repeat are expressed more highly in the parental line LD and progeny 11272 than in line LDX or 11268. These results are consistent with the *Rhg1* copy number estimates. RNA-sequencing is reported in reads per kilobase per million reads and normalized to expression from line LDX01-1-165.



**Figure S4. Soybean lines estimated to contain 3 copies of *Rhg1* display high levels of cytosine methylation in regulatory regions of genes shown to impact SCN resistance.**

A heatmap depicting cytosine methylation levels in 150bp bins at *Rhg1* shows high levels of (A) CG methylation in line LD, estimated to contain 3 copies of *Rhg1*, relative to line LDX, estimated to contain a single copy of *Rhg1*. Their progeny were also assayed for cytosine methylation and progeny estimated to contain 3 copies of *Rhg1* (shown in blue) have a similarly high level of CG methylation compared to single copy progeny (shown in green). The progeny were selected in the F3 generation and likely contain lines with heterozygous *Rhg1* loci (shown in black). (B) Cytosine methylation was analyzed as in (A) but in the sequence context of CHH, where H is any nucleotide A, T, or C. Chromosome positions are shown for reference. Bins that did not have cytosine sequence data, either because of low coverage or because no cytosine exist

in that sequence context within the bin, are shown as black. Gene models above each graph are shown in scale to chromosome 18 for reference.

## 3.10 Supplemental Tables

**Table S1. Summary statistics for SoyNAM whole genome sequencing**

<b>Genotype</b>	<b>Sequence (Mb)</b>	<b>Average Coverage</b>	<b>Reads</b>	<b>Quality Score</b>	<b>Read Length (bp)</b>
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

Data are listed for each genotype and summarized as the total amount of sequence generated in megabases (Mb). Average Coverage is the genome wide average sequence coverage. Reads corresponds to the number of reads generated and the Quality Score is the average Phred based quality score for the total reads. All sequences were generated from a short insert library with paired-end sequencing of 151 bases.

**Table S1.**

**Table S2. Summary statistics for Hg-Type Test whole genome sequencing**

<b>Genotype</b>	<b>Sequencing Library</b>	<b>Sequence (Mb)</b>	<b>Average Coverage</b>	<b>Reads</b>	<b>Quality Score</b>	<b>Read Length (bp)</b>
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

Data are listed for each genotype, separated into the sequencing for each library type. The total amount of sequence generated in megabases (Mb) and Average Coverage are presented genome wide. Reads corresponds to the number of reads generated and Quality Score is the average Phred based quality score for the total reads. All sequences had a read length of 101 bases.

**Table S2.**

**Table S3. Estimated *Rhg1* copy number for SoyNAM lines using rapid mapping**

<b>Genotype</b>	<b>Copy Number Estimates</b>	
	<b>Chromosome 18 (<i>Rhg1</i>)</b>	<b>Chromosome 11 (paralog)</b>
4J105-34	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
PI574486	1.3 ± 0.2	
LG05-4317	1.3 ± 0.2	
LG97-7012	1.2 ± 0.1	
LG04-4717	1.1 ± 0.6	
LG98-1605	1.1 ± 0.4	
PI427136	1.1 ± 0.3	
PI404188A	1.1 ± 0.3	
LG90-2550	1.1 ± 0.3	
U03-100612	1.1 ± 0.2	
PI398881	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	
PI518751	1.0 ± 0.3	
LG92-1255	1.0 ± 0.3	
HS6-3976	1.0 ± 0.3	
Prohio	1.0 ± 0.2	
PI561370	1.0 ± 0.2	
PI507681B	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	

Short reads from whole genome sequencing were aligned to a portion of the reference genome to rapidly estimate *Rhg1* copy number (chromosome 18), results are listed by

*Table S3 cont'd*

genotype. 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. Similar analysis was performed for the paralogous sequence on chromosome 11, and did not significantly deviate from a single copy.

**Table S3.**

**Table S4. Sequenced cDNA products confirms the expression of multiple alleles of *Glyma18g02590* in the different multi-copy *Rhg1* classes**

<b>Genotype</b>	<b>Identified Allele</b>	<b>Sequenced cDNA products</b>
PI 88788	High copy allele	26
	Low copy allele	0
	Splice isoform	0
	Williams-type	2
PI 209332	High copy allele	8
	Low copy allele	0
	Splice isoform	0
	Williams-type	0
Cloud	High copy allele	7
	Low copy allele	0
	Splice isoform	0
	Williams-type	1
Peking	High copy allele	0
	Low copy allele	8
	Splice isoform	1
	Williams-type	0
PI 90763	High copy allele	0
	Low copy allele	6
	Splice isoform	3
	Williams-type	0
PI 89772	High copy allele	0
	Low copy allele	6
	Splice isoform	1
	Williams-type	0
PI 437654	High copy allele	0
	Low copy allele	8
	Splice isoform	3
	Williams-type	0
Williams 82	High copy allele	0
	Low copy allele	0
	Splice isoform	0
	Williams-type	6

Products from cloning cDNA products are listed by genotype. The identified allele corresponds to different alleles of *Glyma18g02590* from the Hg Type Test lines and Williams 82. A zero indicates the transcript was not observed. A splice isoform was detected in all low-copy genomes not detected in Williams 82 or the high-copy genomes.

**Table S4.**

**Table S5. Amino acid polymorphisms for *Rhg1* paralogous genes encoded on chromosome 11.**

Position (bp)	Peking	PI 90763	PI 89772	PI 437654	LD01- 5907	Cloud	LG05- 4292	Maverick
<b><i>Glyma11g35840</i></b> (Gm18.2570.Paralog)								
<b><i>Glyma11g35830</i></b> (Gm18.2580.Paralog)								
<b><i>Glyma11g35820</i></b> (Gm18.2590.Paralog)								
37418427	Isoform	Isoform	Isoform	Isoform	Isoform	Isoform	Isoform	Isoform
37418685	A179T	A179T	A179T	A179T	A179T	A179T	A179T	
<b><i>Glyma11g35810</i></b> (Gm18.2600.Paralog)								
37413493	R320Q	R320Q	R320Q	R320Q	R320Q	R320Q	R320Q	R320Q
37413566	A296T	A296T	A296T	A296T	A296T	A296T	A296T	A296T
<b><i>Glyma11g35800</i></b> (Gm18.2610.Paralog)								
<b><i>Glyma11g35790</i></b> (Gm18.2620.Paralog)								

Position : Chromosome 11 base-pair position relative to Williams 82 reference genome, with gene name (and chromosome18 paralog) above relevant bp positions. Of the six genes analyzed, four did not contain polymorphisms relative to Williams 82 as indicated. Isoform: An mRNA splice isoform predicted to be caused by a SNP as reported (Matsye et al., 2012). Amino acid polymorphisms are reported as the amino acid present in Williams 82, the amino acid position, and the resulting new amino acid discovered. Genotypes not listed did not show polymorphic amino acid sequence for the genes analyzed.

**Table S5.**

**Table S6. Sequence frequencies at DNA variant positions across the *Rhg1* repeat indicates varying sequence content between copies.**

Position	Genotypes			Average Frequency
	LD00-3309	PI 209332	Cloud	
1633532	1.00	1.00	1.00	1.00
1633629	1.00	1.00	1.00	1.00
1633700	1.00	1.00	1.00	1.00
1633840	1.00	0.99	1.00	1.00
1633930	1.00	1.00	1.00	1.00
1634533	1.00	0.90	1.00	0.97
1634534	1.00	0.92	1.00	0.97
1634535	1.00	0.92	1.00	0.97
1634536	1.00	0.92	1.00	0.97
1634610	1.00	1.00	1.00	1.00
1634620	1.00	1.00	1.00	1.00
1634626	1.00	1.00	1.00	1.00
1634635	1.00	1.00	1.00	1.00
1634643	1.00	1.00	1.00	1.00
1634714	1.00	1.00	1.00	1.00
1634856	1.00	1.00	1.00	1.00
1635001	1.00	1.00	1.00	1.00
1635014	1.00	1.00	1.00	1.00
1635093	1.00	1.00	1.00	1.00
1635120	1.00	1.00	1.00	1.00
1635364	1.00	1.00	1.00	1.00
1635912	1.00	1.00	1.00	1.00
1636766	1.00	1.00	1.00	1.00
1639354	0.96	0.99	1.00	0.98
1640056	1.00	1.00	1.00	1.00
1640137	1.00	1.00	1.00	1.00
1640151	1.00	1.00	1.00	1.00
1640292	1.00	1.00	1.00	1.00
1640480	0.92	0.97	1.00	0.96
1640581	0.88	0.90	0.81	0.86
1640675	0.86	0.90	0.75	0.84
1641208	0.88	0.87	0.88	0.88
1641800	0.89	0.87	0.88	0.88
1642266	0.79	0.92	0.81	0.84
1642672	0.91	0.92	0.84	0.89
1642762	0.94	0.95	0.91	0.93
1642848	0.87	0.90	0.81	0.86
1643208	0.91	0.88	0.87	0.89
1643849	0.89	0.89	0.87	0.89
1644089	0.91	0.87	0.83	0.87
1644385	0.90	0.96	0.90	0.92
1644493	0.88	0.93	0.88	0.90
1644525	0.89	0.89	0.84	0.87
1644577	0.85	0.87	0.84	0.85
1644965	0.85	0.87	0.81	0.84
1644968	0.84	0.88	0.81	0.84
1644972	0.87	0.92	0.87	0.89
1644974	0.84	0.88	0.83	0.85

1645218	0.84	0.89	0.77	0.83
1645437	0.86	0.93	0.92	0.90
1645745	0.90	0.91	0.86	0.89
1645759	0.88	0.88	0.85	0.87
1645811	0.92	0.87	0.84	0.88
1645908	0.88	0.89	0.87	0.88
1645914	0.93	0.92	0.93	0.93
1646130	0.81	0.70	0.54	0.68
1646138	0.77	0.62	0.51	0.63
1646145	0.76	0.61	0.50	0.62
1646211	0.88	0.85	0.82	0.85
1646226	0.91	0.86	0.83	0.87
1648850	0.93	0.94	0.85	0.91
1649069	0.96	0.90	0.88	0.91
1649212	0.91	0.89	0.86	0.88
1649293	0.96	0.91	0.87	0.92
1649328	0.94	0.92	0.87	0.91
1649335	1.00	0.93	0.81	0.91
1649371	0.92	0.89	0.85	0.88
1649385	0.90	0.90	0.83	0.88
1649553	0.97	0.97	0.92	0.95
1649630	0.84	0.93	0.85	0.87
1649892	0.92	0.91	0.85	0.89
1649934	0.91	0.87	0.88	0.88
1650106	0.93	0.94	0.93	0.93
1650140	0.90	0.91	0.90	0.90
1650201	0.89	0.92	0.89	0.90
1650310	0.98	0.95	0.92	0.95
1650501	0.89	0.86	0.88	0.88
1650590	0.87	0.93	0.98	0.93
1650772	0.90	0.88	0.86	0.88
1651003	0.95	1.00	1.00	0.98
1651056	0.93	0.93	0.92	0.92
1651602	0.95	0.91	0.90	0.92
1652723	0.92	0.93	1.00	0.95
1653661	0.82	0.90	0.87	0.86
1654230	0.87	0.96	0.82	0.88
1654282	0.94	0.92	0.80	0.89
1655099	0.88	0.87	0.72	0.82
1655195	0.86	0.92	0.86	0.88
1655348	0.94	0.99	0.89	0.94
1655353	0.89	0.92	0.85	0.89
1655408	0.95	0.90	0.87	0.91
1655500	0.94	0.95	0.90	0.93
1655564	0.96	0.94	0.82	0.91
1655585	0.93	0.87	0.82	0.87
1655836	0.88	0.89	0.84	0.87
1656044	0.90	0.88	0.80	0.86
1656178	0.88	0.94	0.85	0.89
1656263	0.88	0.89	0.82	0.86
1656394	0.88	0.87	0.83	0.86
1656417	0.82	0.82	0.71	0.79
1656462	0.81	0.84	0.85	0.83
1656633	0.88	0.94	0.94	0.92

1656719	0.91	0.97	0.79	0.89
1656769	0.88	0.91	0.83	0.87
1656898	0.89	0.91	0.81	0.87
1656979	0.92	0.95	0.84	0.90
1657025	0.42	0.54	0.51	0.49
1657162	0.85	0.91	0.91	0.89
1657183	0.86	0.93	0.90	0.90
1657307	0.91	0.86	0.87	0.88
1657506	0.91	0.88	0.88	0.89
1657803	0.87	0.88	0.74	0.83
1657807	0.83	0.76	0.68	0.76
1657815	0.84	0.76	0.73	0.78
1657816	0.84	0.74	0.70	0.76
1658170	0.90	0.93	0.87	0.90
1658284	0.83	0.91	0.81	0.85
1658617	0.86	0.88	0.91	0.88
1658735	0.90	0.91	0.88	0.90
1659502	0.91	0.95	0.93	0.93
1659777	0.88	0.92	0.86	0.89
1659829	0.83	0.89	0.83	0.85
1659914	0.91	0.88	0.87	0.89
1659945	0.91	0.83	0.88	0.87
1660067	0.91	0.88	0.82	0.87
1660183	0.89	0.91	0.89	0.90
1660790	1.00	1.00	1.00	1.00
1661155	1.00	1.00	1.00	1.00
1661264	0.87	0.72	0.62	0.74
1661293	0.87	0.78	0.74	0.80
1661406	0.89	0.92	0.86	0.89
1661428	0.91	0.90	0.86	0.89
1661460	0.95	0.93	0.84	0.91
1662031	0.96	0.91	0.86	0.91
1662115	1.00	1.00	1.00	1.00
1662177	0.88	0.87	0.86	0.87
1662656	0.84	0.94	0.89	0.89
1662666	0.80	0.91	0.89	0.87
1662682	0.81	0.93	0.91	0.88
1662714	0.86	0.95	0.91	0.91
1662734	1.00	1.00	1.00	1.00
1662810	0.82	0.85	0.85	0.84
1662851	1.00	1.00	1.00	1.00
1662946	0.83	0.83	0.80	0.82
1662953	0.71	0.83	0.79	0.78
1663007	0.63	0.90	0.86	0.80
1663014	0.63	0.90	0.82	0.78
1663032	0.60	0.90	0.85	0.79
1663064	0.57	0.90	0.86	0.78
1663114	0.69	0.94	0.86	0.83
1663133	0.65	0.92	0.78	0.78
1663148	0.66	0.92	0.76	0.78
1663225	1.00	1.00	1.00	1.00
1663250	0.70	0.94	0.78	0.81

*Table S6 cont'd*

Position: Base pair position on chromosome 18 of Williams 82 corresponding to a DNA variant position. The variant allele frequency is reported below each genotype as the sum of 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. Average Frequency: The average frequency at a given variant site computed from the three genomes.

**Table S6.**

### 3.11 References

- Aitman TJ, Dong R, Vyse TJ, Norsworthy PJ, Johnson MD, Smith J, Mangion J, Robertson-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
- Arelli APR, Webb DM** (1996) Molecular genetic diversity among soybean plant introductions with resistance to *Heterodera glycines*. *Current Science* **71**: 230-233
- Barnard RJ, Morgan A, Burgoyne RD** (1996) Domains of alpha-SNAP required for the stimulation of exocytosis and for N-ethylmaleimide-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
- Bass C, Field LM** (2011) Gene amplification and insecticide resistance. *Pest Manag Sci* **67**: 886-890
- 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
- Bryant D, Moulton V** (2004) Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol* **21**: 255-265
- Caldwell 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 Illinois, Urbana, IL
- Cary T, Diers BW** (2010) 2009 Northern regional soybean cyst nematode tests. *In*. University of Illinois, Urbana, IL
- Cary T, Diers BW** (2011) 2010 Northern regional soybean cyst nematode tests. *In*. University of Illinois, Urbana, IL

- Cary T, Diers BW** (2012) 2011 Northern regional soybean cyst nematode tests. *In*. University of Illinois, Urbana, IL
- Chang AYF, Liao BY** (2012) DNA Methylation Rebalances Gene Dosage after Mammalian Gene Duplications. *Molecular biology and evolution* **29**: 133-144
- Chen ZJ** (2007) Genetic and Epigenetic Mechanisms for Gene Expression and Phenotypic Variation in Plant Polyploids. *Annu Rev Plant Biol* **58**: 377-406
- 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
- 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
- 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
- Cregan P, Diers B** (unpublished) Whole genome sequence analysis of SoyNAM: soybean nested association mapping population. Unpublished
- Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Graf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S, Langerod A, Green A, Provenzano E, Wishart G, Pinder S, Watson P, Markowitz F, Murphy L, Ellis I, Purushotham A, Borresen-Dale AL, Brenton JD, Tavare S, Caldas C, Aparicio S** (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**: 346-352
- 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
- DeBolt S** (2010) Copy Number Variation Shapes Genome Diversity in *Arabidopsis* Over Immediate Family Generational Scales. *Genome Biology and Evolution* **2**: 441-453
- 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
- Downen RH, Pelizzola M, Schmitz RJ, Lister R, Downen 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
- 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
- Feuk L, Carson AR, Scherer SW** (2006) Structural variation in the human genome. *Nature Reviews Genetics* **7**: 85-97
- Flagel LE, Wendel JF** (2009) Gene duplication and evolutionary novelty in plants. *New Phytologist* **183**: 557-564
- Gohlke J, Scholz C-J, Kneitz S, Weber D, Fuchs J, Hedrich R, Deeken R** (2013) DNA Methylation Mediated Control of Gene Expression Is Critical for Development of Crown Gall Tumors. *PLoS Genet* **9**: e1003267
- Heinberg A, Siu E, Stern C, Lawrence EA, Ferdig MT, Deitsch KW, Kirkman LA** (2013) Direct evidence for the adaptive role of copy number variation on antifolate susceptibility in *Plasmodium falciparum*. *Mol Microbiol* **88**: 702-712
- Hernando-Herraez I, Prado-Martinez J, Garg P, Fernandez-Callejo M, Heyn H, Hvilson C, Navarro A, Esteller M, Sharp AJ, Marques-Bonet T** (2013) Dynamics of DNA Methylation in Recent Human and Great Ape Evolution. *PLoS Genet* **9**: e1003763
- Huson DH, Bryant D** (2006) Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**: 254-267
- Innan H, Kondrashov F** (2010) The evolution of gene duplications: classifying and distinguishing between models. *Nature Reviews Genetics* **11**: 97-108

- Kenrick P, Crane PR** (1997) The origin and early evolution of plants on land. *Nature* **389**: 33-39
- 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
- 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
- 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, Rogozin IB, Wolf YI, Koonin EV** (2002) Selection in the evolution of gene duplications. *Genome Biol* **3**: 9
- 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

- 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
- 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
- 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
- Maron LG, Guimaraes CT, Kirst M, Albert PS, Birchler JA, Bradbury PJ, Buckler ES, Coluccio AE, Danilova TV, Kudrna D, Magalhaes JV, Pinerros 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, 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
- 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
- Mikel MA, Diers BW, Nelson RL, Smith HH** (2010) Genetic Diversity and Agronomic Improvement of North American Soybean Germplasm. *Crop Science* **50**: 1219-1229

- 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
- 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 virulence 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
- 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
- Olsen KM, Wendel JF** (2013) A bountiful harvest: genomic insights into crop domestication phenotypes. *Annu Rev Plant Biol* **64**: 47-70
- Perry GH, Yang F, Marques-Bonet T, Murphy C, Fitzgerald T, Lee AS, Hyland C, Stone AC, Hurler 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
- Qian WF, Liao BY, Chang AYP, Zhang JZ** (2010) Maintenance of duplicate genes and their functional redundancy by reduced expression. *Trends in Genetics* **26**: 425-430
- Quinlan AR, Hall IM** (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842
- 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
- Roberts I, Carter SA, Scarpini CG, Karagavriilidou K, Barna JC, Calleja M, Coleman N** (2012) A high-throughput computational framework for identifying significant copy

number aberrations from array comparative genomic hybridisation data. *Adv Bioinformatics* **2012**: 876976

**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

**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*

**Schmitz RJ, Schultz MD, Urich MA, Nery JR, Pelizzola M, Libiger O, Alix A, McCosh RB, Chen H, Schork NJ, Ecker JR** (2013) Patterns of population epigenomic diversity. *Nature* **495**: 193-198

**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 Academic Publishers

**Stebbing J, Filipovic A, Lit LC, Blighe K, Grothey A, Xu Y, Miki Y, Chow LW, Coombes RC, Sasano H, Shaw JA, Giamas G** (2013) LMTK3 is implicated in endocrine resistance via multiple signaling pathways. *Oncogene* **32**: 3371-3380

**Stemans 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

**Sutherland E, Coe L, Raleigh EA** (1992) McrBC: a multisubunit GTP-dependent restriction endonuclease. *J Mol Biol* **225**: 327-348

**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

**Tang YC, Amon A** (2013) Gene copy-number alterations: a cost-benefit analysis. *Cell* **152**: 394-405

**Team RDC** (2009) R: A language and environment for statistical computing. *In*. R foundation for Statistical Computing

**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

- Tylka GL, Gebhart GD, Marrett CC, Mullaney MP** (2012) Evaluation of soybean varieties resistant to soybean cyst nematode in Iowa in 2012. *In*. Iowa State University Extension and Outreach
- USDA A, National Genetic Resources Program** (2014) Germplasm resources information network (GRIN). *In*. National Germplasm Resources Laboratory, Beltsville, Maryland
- 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
- Webb DM** (2012) Quantitative trait loci associated with soybean cyst nematode resistance and uses thereof. *In* USPaT Office, ed, Vol US 20120278953 A1, USA
- 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
- Wrather JA, Koenning SR** (2009) Effects of Diseases on Soybean Yields in the United States 1996 to 2007. *In* Plant Health Progress, Ed 2009
- Wu HJ, Zhang ZH, Wang JY, Oh DH, Dassanayake M, Liu BH, Huang QF, Sun HX, Xia R, Wu YR, Wang YN, Yang Z, Liu Y, Zhang WK, Zhang HW, Chu JF, Yan CY, Fang S, Zhang JS, Wang YQ, Zhang FX, Wang GD, Lee SY, Cheeseman JM, Yang BC, Li B, Min JM, Yang LF, Wang J, Chu CC, Chen SY, Bohnert HJ, Zhu JK, Wang XJ, Xie Q** (2012) Insights into salt tolerance from the genome of *Thellungiella salsuginea*. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 12219-12224
- 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
- Young LD** (1996) Yield loss in soybean caused by *Heterodera glycines*. *Journal of Nematology* **28**: 604-607
- Yu A, Lepere G, Jay F, Wang JY, Bapaume L, Wang Y, Abraham AL, Penterman J, Fischer RL, Voinnet O, Navarro L** (2013) Dynamics and biological relevance of DNA demethylation in *Arabidopsis* antibacterial defense. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 2389-2394
- Ziller MJ, Gu H, Muller F, Donaghey J, Tsai LTY, Kohlbacher O, De Jager PL, Rosen ED, Bennett DA, Bernstein BE, Gnirke A, Meissner A** (2013) Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500**: 477-481

## **Chapter 4: Disease resistance through impairment of $\alpha$ -SNAP/NSF interaction and vesicular trafficking by soybean *Rhg1***

This chapter was previously published as:

Bayless AM, Smith JM, Song J, McMinn PH, Teillet A, August BK, Bent AF (2016) Disease resistance through impairment of alpha-SNAP-NSF interaction and vesicular trafficking by soybean *Rhg1*. Proc Natl Acad Sci U S A 113: E7375-E7382

Contributions: I conceived of this project with Andrew Bent. I performed the majority of the described experiments and wrote the manuscript with Andrew Bent. John Smith performed the soybean co-IP assay, sec-GFP trafficking, electromicroscopy and syncytial Western blot and assisted with manuscript writing. Junqi Song performed the *N. benthamiana* co-IP assays.

#### 4.1 Abstract

$\alpha$ -SNAP and NSF proteins are conserved across eukaryotes and sustain cellular vesicle trafficking by mediating disassembly and reuse of SNARE protein complexes, which facilitate fusion of vesicles to target membranes. However, certain haplotypes of the *Rhg1* locus of soybean possess multiple repeat copies of an  $\alpha$ -SNAP gene (*Glyma.18g022500*) that encodes atypical amino acids at a highly conserved functional site. These *Rhg1* loci mediate resistance to soybean cyst nematode (SCN; *Heterodera glycines*), the most economically damaging pathogen of soybeans worldwide. *Rhg1* is widely utilized in agriculture but the mechanisms of *Rhg1* disease resistance have remained unclear. In the present study we found that the resistance-type *Rhg1*  $\alpha$ -SNAP is defective in interaction with NSF. Elevated *in planta* expression of resistance-type *Rhg1*  $\alpha$ -SNAPs depleted the abundance of SNARE-recycling 20S complexes, disrupted vesicle trafficking, induced elevated abundance of NSF, and caused cytotoxicity. Soybean, due to ancient genome duplication events, carries other loci that encode canonical (wild-type)  $\alpha$ -SNAPs. Expression of these  $\alpha$ -SNAPs counteracted the cytotoxicity of resistance-type *Rhg1*  $\alpha$ -SNAPs. For successful growth and reproduction, SCN dramatically reprograms a set of plant root cells and must sustain this sedentary feeding site for two to four weeks. Immunoblots and electron microscopy immunolocalization revealed that resistance-type  $\alpha$ -SNAPs specifically hyperaccumulate relative to wild-type  $\alpha$ -SNAPs at the nematode feeding site, promoting the demise of this biotrophic interface. The paradigm of disease resistance through a dysfunctional variant of an essential gene was first established for humans and malaria, and may be applicable to other plant-pathogen interactions.

## 4.2 Significance Statement

The *Rhg1* resistance locus of soybean helps control one of the most damaging diseases in world agriculture. We found that *Rhg1*-mediated resistance functions by an unusual mechanism. Resistant soybeans carry a dysfunctional variant of the housekeeping protein  $\alpha$ -SNAP (Soluble NSF Attachment Protein). *Rhg1* resistance-type  $\alpha$ -SNAPs interact poorly with NSF (N-ethylmaleimide-Sensitive Factor) and disrupt vesicle trafficking. High levels of resistance-type  $\alpha$ -SNAPs interfere with wild-type  $\alpha$ -SNAP activities, but are functionally balanced in most tissues by sufficient wild-type  $\alpha$ -SNAP levels. However, the biotrophic plant-pathogen interface is disabled by localized hyperaccumulation of resistance-type  $\alpha$ -SNAPs. This study suggests a paradigm of resistance conferred by a dysfunctional version of a core cellular housekeeping protein.

### 4.3 Introduction

A dynamic endomembrane system is a universal trait of eukaryotic cells that enables the transfer of vesicular cargoes throughout the cell and with the cell exterior (Jahn and Scheller, 2006). Vesicle trafficking has been most deeply studied in yeast and neuronal synapses, but is understood in detail for numerous biological systems including immunity and host-pathogen interactions (Collins et al., 2003; Wickner and Schekman, 2008; Asrat et al., 2014; Inada and Ueda, 2014). Host and pathogen proteins can intervene to alter the course of this traffic to the benefit of the host or the pathogen (Hoefle and Huckelhoven, 2008; Uemura et al., 2012). The soybean (*Glycine max*) *Rhg1* (Resistance to *Heterodera glycines* 1) locus, one of the most economically important disease resistance loci of any major food crop, carries multiple repeat copies of a gene encoding the major vesicular trafficking chaperone  $\alpha$ -SNAP (alpha-Soluble NSF Attachment Protein) (Jahn and Scheller, 2006; Cook et al., 2012). The discovery that the *Rhg1*  $\alpha$ -SNAPs carry non-consensus amino acids at widely conserved C-terminal positions of known importance was intriguing, but a mechanism by which these  $\alpha$ -SNAPs contribute to *Rhg1*-mediated soybean cyst nematode (*Heterodera glycines*) resistance was not known (Barnard et al., 1997; Cook et al., 2012; Cook et al., 2014).

SNARE (Soluble NSF Attachment Protein Receptors) proteins mediate vesicle fusion (Jahn and Scheller, 2006). Eukaryote genomes can encode over one hundred different SNARE proteins, with various SNARE subsets generally residing at specific compartments (Jahn and Scheller, 2006). Cognate SNAREs on separate membranes promote fusion by bundling together and forming highly stable SNARE complexes that pull the respective membranes together. SNAREs alone can mediate vesicle fusion *in vitro* without external energy inputs, but the cis-SNARE complexes formed after fusion must be separated back into free acceptor SNAREs to

participate in subsequent fusion events (Jahn and Scheller, 2006).  $\alpha$ -SNAP, which is typically encoded by a single gene in animal genomes, binds diverse SNARE complexes and stimulates their disassembly by recruiting and activating NSF (N-ethylmaleimide- Sensitive Factor) (Jahn and Scheller, 2006; Vivona et al., 2013). SNARE complex disassembly by  $\alpha$ -SNAP and NSF is essential for vesicular trafficking and as such, has been studied in considerable detail. X-ray crystallography, single-molecule fluorescence spectroscopy and cryo-EM have provided high-resolution structural insights into the dynamics of SNARE/ $\alpha$ -SNAP/NSF interactions (Ryu et al., 2015; Zhao et al., 2015; Zick et al., 2015). Multiple  $\alpha$ -SNAPs stimulate disassembly of one SNARE bundle in a 20S supercomplex that includes a hexameric ring of six NSF proteins, which couple ATP-hydrolysis to force-generating conformational changes.

Cyst nematodes are highly adapted obligate parasites of plant roots and cause substantial damage to world food crops including wheat, soybean and potato (Gheysen and Mitchum, 2011). Soybean cyst nematode (SCN) is responsible for the greatest yield loss in the U.S. of any soybean disease and is a major constraint on soybean production worldwide (Niblack et al., 2006). After penetrating the root and migrating to the root vascular bundle, SCN secrete plant-bioactive effector proteins and other molecules through their stylet – a protrusible mouthpiece that also mediates nematode feeding on plant cells. SCN effectors collectively subdue host defenses and reprogram root cells to fuse and form a metabolically hyperactive syncytium (nematode feeding site) (Davis et al., 2008; Gheysen and Mitchum, 2011). Syncytium formation is a complex process involving plant cell wall dissolution, endoreduplication, cell-cell fusion and membrane reorganization, with the eventual incorporation of over one hundred host root cells into one large multinucleate cell (Niblack et al., 2006; Gheysen and Mitchum, 2011; Kyndt et al., 2013). Because egg-filled SCN cysts can persist in fields for many years and nematicides are

often costly and environmentally damaging, the two core SCN control strategies are crop rotation to reduce inoculum load, and use of SCN-resistant soybean varieties.

The soybean *Rhg1* quantitative trait locus provides the strongest known SCN resistance (Concibido et al., 2004; Donald et al., 2006). Recently, the *Rhg1* locus was molecularly isolated and characterized (Cook et al., 2012). Surprisingly, the *Rhg1* locus is a repeated block of four disparate genes that do not resemble previously known plant disease resistance mediators. Gene-silencing and gene complementation experiments demonstrated contributions to SCN resistance for three of the four tightly linked genes on the *Rhg1* repeat: *Glyma.18g022400* (encoding a putative amino acid permease, formerly *Glyma18g02580*), *Glyma.18g022500* (encoding a predicted  $\alpha$ -SNAP, formerly *Glyma18g02590*) and *Glyma.18g022700* (a predicted wound-inducible protein, formerly *Glyma18g02610*) (Cook et al., 2012). Expressing any single gene found within repeated *Rhg1* blocks, including a unique polymorphic  $\alpha$ -SNAP, did not elevate SCN resistance; simultaneous expression of this polymorphic  $\alpha$ -SNAP with the other *Rhg1* block encoded genes enhanced SCN resistance (8). The ~30 kb *Rhg1* segment is present in a single copy in SCN-susceptible soybean varieties, but multiple direct repeat copies are present in SCN-resistant varieties (Cook et al., 2012). Two distinct classes of resistance-encoding *Rhg1* haplotypes have been identified, low-copy (3 copies or less) and high-copy (>4 copies); the low- and high-copy *Rhg1* haplotypes each encode distinct polymorphic  $\alpha$ -SNAPs (Fig. 1A) (Cook et al., 2014; Lee et al., 2015). No amino acid polymorphisms are predicted in the *Glyma.18g022400* or *Glyma.18g022700* products from SCN-susceptible as opposed to SCN-resistant *Rhg1* haplotypes, but *Rhg1* copy number expansion constitutively elevates the transcript levels of these genes in SCN-resistant plants (Cook et al., 2014). The polymorphisms in the *Rhg1* *Glyma.18g022500*-encoded  $\alpha$ -SNAP are at the highly conserved C-terminus, which in mammal

and yeast systems directly contacts NSF and is required for activation of SNARE disassembly (Barnard et al., 1997; Cook et al., 2012; Cook et al., 2014; Zick et al., 2015).

For over 30 years the soybean industry has relied on extensive use of *Rhg1* from a single source, PI88788 (Colgrove and Niblack, 2008). Field SCN populations evolve slowly but are increasingly exhibiting partial virulence on plants expressing PI88788-derived *Rhg1* (Lambert et al., 2005; Niblack et al., 2008). Understanding the molecular mechanisms of *Rhg1*-mediated SCN resistance may allow quantitative improvements to *Rhg1* resistance through allele diversification, the generation of synthetic improved resistance, and/or transfer of the widely successful *Rhg1*-mediated resistance mechanism to other crops such as wheat or potato. In this study we utilized *in vitro* and *in planta* methods to functionally characterize the *Rhg1*-encoded  $\alpha$ -SNAPs. We discovered the unusual presence of a stably inherited  $\alpha$ -SNAP that is toxic to normal  $\alpha$ -SNAP/NSF interactions and vesicular trafficking, yet is beneficial during the *Rhg1*-mediated SCN resistance response of soybean.

## 4.4 Results

### ***Rhg1* Resistance-Type $\alpha$ -SNAPs that are Polymorphic at Conserved Residues are Impaired in NSF Interactions**

The C-terminal six amino acid residues of  $\alpha$ -SNAPs are very highly conserved across eukaryotes, with three or four acidic residues followed by the near-universal penultimate leucine (Fig. S1A). Most soybeans are susceptible to SCN and their single-copy *Rhg1* locus  $\alpha$ -SNAP matches this consensus, but the SCN resistance-conferring high-copy or low-copy *Rhg1* loci encode multiple copies of  $\alpha$ -SNAPs that diverge at these sites and an upstream residue (Fig. 1A) (Cook et al., 2014; Lee et al., 2015). Since electrostatic contacts between the NSF N domain and the acidic residues at the  $\alpha$ -SNAP C-terminus are reported in animal systems (Zhao et al., 2015), we examined NSF binding by *Rhg1*  $\alpha$ -SNAPs. The reference Williams 82 soybean genome encodes two NSF proteins, *Glyma.07G195900* (NSF<sub>Ch7</sub>) and *Glyma.13G180100* (NSF<sub>Ch13</sub>), which are 98% identical. For *in vitro* binding studies we generated recombinant NSF<sub>Ch7</sub> and NSF<sub>Ch13</sub> proteins, as well as recombinant *Rhg1*  $\alpha$ -SNAP proteins of the high-copy type (PI 88788-type) and low-copy type (Peking-type), designated as  $\alpha$ -SNAP<sub>*Rhg1*HC</sub> and  $\alpha$ -SNAP<sub>*Rhg1*LC</sub>, and the SCN-susceptible Williams 82 wild-type  $\alpha$ -SNAP ( $\alpha$ -SNAP<sub>*Rhg1*WT</sub>). *In vitro* NSF binding assays were performed essentially as in (Barnard et al., 1996). We observed that NSF<sub>Ch7</sub> or NSF<sub>Ch13</sub> binding to either  $\alpha$ -SNAP<sub>*Rhg1*HC</sub> or  $\alpha$ -SNAP<sub>*Rhg1*LC</sub> was reduced ~60-70% compared to  $\alpha$ -SNAP<sub>*Rhg1*WT</sub> (Fig. 1B,C, S1B,C). In soybean, we have detected an alternatively spliced transcript for the low copy  $\alpha$ -SNAP ( $\alpha$ -SNAP<sub>*Rhg1*LC<sub>Splice</sub></sub>), representing ~20% of total  $\alpha$ -SNAP<sub>*Rhg1*LC</sub> transcripts (Fig. S1D)(Cook et al., 2014). The  $\alpha$ -SNAP encoded by the  $\alpha$ -SNAP<sub>*Rhg1*LC<sub>Splice</sub></sub> transcript, which retains the same C-terminus but removes residues 209–221 (Fig. 1A), also bound NSF poorly (Fig. S1E). The requirement of the soybean  $\alpha$ -SNAP C-

terminus for NSF binding was examined by truncating the final 10 C-terminal residues of  $\alpha$ -SNAP<sub>Rhg1</sub>WT ( $\alpha$ -SNAP<sub>Rhg1</sub>WT(-10)); little to no binding of NSF<sub>Ch7</sub> by this protein was observed (Fig. S1E). We examined the conservation of the  $\alpha$ -SNAP C-terminus in NSF binding across distant eukaryotes by testing the binding of Chinese hamster NSF (NSF<sub>CHO</sub>; 45% identity to soybean NSF) with the soybean *Rhg1*  $\alpha$ -SNAPs. Robust binding of NSF<sub>CHO</sub> to  $\alpha$ -SNAP<sub>Rhg1</sub>WT was observed while NSF<sub>CHO</sub> binding to either  $\alpha$ -SNAP<sub>Rhg1</sub>HC or  $\alpha$ -SNAP<sub>Rhg1</sub>LC was reduced >80%, indicating strong conservation of the  $\alpha$ -SNAP C-terminus for NSF interactions (Fig. S1F,G).

Free, unbound  $\alpha$ -SNAP is not reported to establish NSF binding interfaces; rather, NSF is recruited by  $\alpha$ -SNAPs bound to SNAREs or immobilized on a plastic surface (Barnard et al., 1996; Jahn and Scheller, 2006). To confirm reduced NSF interactions with the *Rhg1* resistance-type  $\alpha$ -SNAPs *in planta*, we performed co-immunoprecipitation (co-IP) assays by expressing soybean NSF<sub>Ch7</sub>-HA and GFP- $\alpha$ -SNAP<sub>Rhg1</sub>HC or GFP- $\alpha$ -SNAP<sub>Rhg1</sub>WT in *Nicotiana benthamiana* leaves via agroinfiltration. Similar to our *in vitro* studies, we reproducibly detected substantial decreases in NSF binding to  $\alpha$ -SNAP<sub>Rhg1</sub>HC compared with  $\alpha$ -SNAP<sub>Rhg1</sub>WT (Fig. 1D). Co-IP in transgenic roots of soybean variety Fayette (high copy *Rhg1*) expressing NSF<sub>Ch7</sub>-HA also demonstrated reduced interactions between NSF<sub>Ch7</sub>-HA and endogenous  $\alpha$ -SNAP<sub>Rhg1</sub>HC as compared with endogenous WT- $\alpha$ -SNAPs (Fig. S1H). Detection of  $\alpha$ -SNAP<sub>Rhg1</sub>WT or  $\alpha$ -SNAP<sub>Rhg1</sub>HC was performed using custom antibodies raised against native peptides mapping to the extreme  $\alpha$ -SNAP C-terminus (See Fig. S2A,B,C for custom antibody specificity).

*Cis*-SNARE complexes formed from vesicle fusion events are recycled in a 20S supercomplex of multiple  $\alpha$ -SNAPs interfaced with the NSF hexamer (Jahn and Scheller, 2006;

Zhao et al., 2015). We examined if 20S complex levels in *N. benthamiana* were affected by  $\alpha$ -SNAP<sub>Rhg1</sub>LC expression. Glycerol gradient ultracentrifugation and fractionation of detergent solubilized membrane proteins determined that  $\alpha$ -SNAP<sub>Rhg1</sub>LC decreased the amount of endogenous membrane associated NSF in 20S fractions by >50% (Fig. 1E,F). A greater proportion of NSF was detected in fractions sedimenting below 20S, suggesting 20S complex instability. On the other hand, with  $\alpha$ -SNAP<sub>Rhg1</sub>WT expression the majority of total membrane associated NSF remained in 20S-sedimenting fractions, similar to empty vector controls (Fig. 1E). Fraction identity was confirmed with parallel co-fractionation of protein standards of known sedimentation (Fig. S3). The specificity of custom antibodies raised against NSF-based peptides was confirmed (Fig. S2D). Because multiple  $\alpha$ -SNAPs participate in stimulating SNARE disassembly by NSF in the 20S complex, 20S destabilization is likely to mean fewer and potentially less productive interactions between wild-type  $\alpha$ -SNAPs and NSF. Together, our *in vitro* and *in planta* results suggest that SCN resistance-conferring  $\alpha$ -SNAPs are compromised in promoting NSF function.

### **Resistance-Type $\alpha$ -SNAPs are Cytotoxic at High Doses and Trigger Elevated NSF Abundance**

We observed that expressing either resistance-type *Rhg1*  $\alpha$ -SNAP ( $\alpha$ -SNAP<sub>Rhg1</sub>LC or  $\alpha$ -SNAP<sub>Rhg1</sub>HC) in *N. benthamiana* caused visible chlorosis 3-4 days after agroinfiltration, with extensive cell death occurring 1-3 days later (Fig. 2A). Cell death induced by  $\alpha$ -SNAP<sub>Rhg1</sub>LC was consistently observed to occur 1-2 days earlier than from  $\alpha$ -SNAP<sub>Rhg1</sub>HC. Expressing  $\alpha$ -SNAP<sub>Rhg1</sub>WT or  $\alpha$ -SNAP<sub>Rhg1</sub>LC<sub>splice</sub> did not result in cell death or other macroscopic phenotypes indicative of stress (Fig. 2A). Expression of  $\alpha$ -SNAP<sub>Rhg1</sub>WT or either resistance-type  $\alpha$ -SNAP

was confirmed using custom antibodies (Fig. 2B, S2A,B,C). The  $\alpha$ -SNAP<sub>Rhg1</sub>LC<sub>Splice</sub> protein was not observed to accumulate in *N. benthamiana* or in transgenic soybean roots (Fig. S4A,B). Cytotoxicity of resistance-type  $\alpha$ -SNAP was also observed, but with a delayed onset, when the proteins were expressed in *N. benthamiana* from the native *Rhg1* promoter in the presence of the other *Rhg1* repeat-associated genes (Fig. S5A). Serial two-fold dilutions of  $\alpha$ -SNAP<sub>Rhg1</sub>LC delivery confirmed dose-sensitivity of the observed cytotoxicity (Fig. S5B).

To test the hypothesis that the cytotoxicity of the unusual resistance-type  $\alpha$ -SNAPs may be due to disruption of NSF-dependent processes, we tested if a defective NSF would recapitulate this cytotoxicity. Mutagenizing a conserved glutamate in the NSF D1 domain Walker B motif generates a dominant-negative ATPase-null NSF (Dalal et al., 2004). We assessed the impact of directly blocking NSF ATPase in *N. benthamiana* by generating the analogous mutation in soybean NSF (NSF<sub>Ch7</sub>-E332Q). Expressing NSF<sub>Ch7</sub>-E332Q caused a cytotoxic symptom onset and severity similar to  $\alpha$ -SNAP<sub>Rhg1</sub>LC expression, while expression of wild-type NSF<sub>Ch7</sub> had no effect, similar to empty vector controls (Fig. S6).

A strong increase in abundance of the endogenous *N. benthamiana* NSF protein was consistently detected in leaves expressing cytotoxic  $\alpha$ -SNAPs while expression of  $\alpha$ -SNAP<sub>Rhg1</sub>WT did not affect NSF levels (Fig. 2B, see also Fig. 2F,H). To determine if elevated NSF expression was specific to resistance-type  $\alpha$ -SNAP expression or a hallmark of stressed cells, we treated leaves with 50  $\mu$ M of the herbicide paraquat for 24 hrs and did not observe significant changes in NSF expression (Fig. S7A). No significant changes in NSF expression were observed in transgenic soybean roots expressing resistance-type  $\alpha$ -SNAPs (S7B,C). Nonetheless, modulation of NSF protein levels from disrupting  $\alpha$ -SNAP function is apparently

unreported in other systems and may be a feedback mechanism characteristic to some plants (Zhao et al., 2007; Barszczewski et al., 2008; Naydenov et al., 2012).

### **Resistance-Type $\alpha$ -SNAPs Disrupt Secretion and *trans*-Golgi Network Trafficking**

Regeneration of free acceptor SNAREs via *cis*-SNARE complex disassembly is necessary for ongoing vesicle trafficking. Because resistance-type  $\alpha$ -SNAPs interacted poorly with NSF, we assessed their impacts on exocytic trafficking in *N. benthamiana* using the sec-GFP secreted GFP assay (Batoko et al., 2000). In this assay, if the engineered sec-GFP protein is secreted extracellularly from the ER to the apoplast it fluoresces weakly but if trafficking is disrupted and sec-GFP is retained in the ER-Golgi network, it fluoresces strongly (Batoko et al., 2000). Samples were monitored at 2 and 3 days after agroinfiltration, before the onset of chlorotic leaf symptoms. Resistance-type  $\alpha$ -SNAP co-expression with sec-GFP strongly induced intracellular sec-GFP fluorescence, while  $\alpha$ -SNAP<sub>Rhg1</sub>WT resembled empty-vector controls and did not perturb sec-GFP trafficking, as evidenced by a lack of fluorescence accumulation (Fig. 2C,D). We additionally examined if resistance-type  $\alpha$ -SNAPs affect Golgi network trafficking using the *trans*-Golgi network (TGN)/early endosome marker Syp61-mCherry (Gu and Innes, 2011). In the vast majority of cells, co-expression of  $\alpha$ -SNAP<sub>Rhg1</sub>WT did not substantially alter the punctate vesicle and plasma membrane distribution and abundance of Syp61 fluorescence seen in empty vector controls.  $\alpha$ -SNAP<sub>Rhg1</sub>LC expression, however, shifted the Syp61-mCherry signal to an extensive and diffuse distribution (albeit excluded from chloroplasts, nuclei and vacuoles)(Fig. S8). The sec-GFP and Syp61-mCherry results indicate that high expression of resistance-type  $\alpha$ -SNAPs disrupts exocytosis and normal trafficking through the Golgi.

## Substituting the Penultimate Leucine Modulates Cell Death Progression from Resistance-Type $\alpha$ -SNAPs

The penultimate leucine is seemingly conserved across all other available plant and animal  $\alpha$ -SNAP sequences, yet resistance-type soybean *Rhg1*  $\alpha$ -SNAPs have an isoleucine at this position (Fig. S1A) (Barnard et al., 1997). *In vitro* studies of yeast and animal NSF have demonstrated that this leucine enhances NSF ATPase activity, and that  $\alpha$ -SNAP with an engineered leucine-to-alanine substitution at this position no longer stimulates ATPase activity or SNARE disassembly (Barnard et al., 1997; Zick et al., 2015). We therefore assessed the effects of penultimate leucine substitutions in the *Rhg1*  $\alpha$ -SNAPs. Curiously, no cytotoxic symptoms in *N. benthamiana* were apparent from  $\alpha$ -SNAP<sub>*Rhg1*WT-L288A</sub> or  $\alpha$ -SNAP<sub>*Rhg1*WT-L288I</sub> (Fig. 2E). However, we detected that NSF protein levels were substantially elevated by  $\alpha$ -SNAP<sub>*Rhg1*WT-L288A</sub> and not by  $\alpha$ -SNAP<sub>*Rhg1*WT</sub> or  $\alpha$ -SNAP<sub>*Rhg1*WT-L288I</sub>, further suggesting that elevation of endogenous NSF is due to dysfunctional  $\alpha$ -SNAPs and not cell death (Fig. 2F). In an otherwise wild-type  $\alpha$ -SNAP, absence of the penultimate leucine is sufficient to trigger increases in NSF protein abundance, but not cell death.

In contrast to results with the wild-type  $\alpha$ -SNAP, in  $\alpha$ -SNAP<sub>*Rhg1*LC</sub> (see Fig. 1A), alanine substitution ( $\alpha$ -SNAP<sub>*Rhg1*LC-I289A</sub>) did enhance the progression of chlorosis and cytotoxicity (Fig. 2G). Conversely, placing a penultimate leucine in a resistance-type  $\alpha$ -SNAP ( $\alpha$ -SNAP<sub>*Rhg1*LC-I289L</sub>) modestly reduced toxicity progression compared to unaltered  $\alpha$ -SNAP<sub>*Rhg1*LC</sub> (Fig. 2G). All  $\alpha$ -SNAP<sub>*Rhg1*LC</sub> substitutions eventually resulted in chlorosis and cell death with large increases in NSF production (Fig. 2H). Similar results were observed with  $\alpha$ -SNAP<sub>*Rhg1*HC</sub> penultimate substitutions (Fig. S9A). Expressing the C-terminally truncated  $\alpha$ -SNAP<sub>*Rhg1*WT(-10)</sub>, which did not strongly bind NSF *in vitro*, elicited strong cytotoxic effects, similar to

resistance-type  $\alpha$ -SNAPs (Fig. S9B). Overall, these results indicate that substitution of the penultimate leucine for isoleucine contributes to the *in planta* cytotoxicity of resistance-type *Rhg1*  $\alpha$ -SNAPs, but the other C-terminal residue changes also contribute to the full effect. The results further indicate that presence of a penultimate isoleucine, compared with a more extreme change (such as leucine-to-alanine), apparently mutes the severity of the resistance-type *Rhg1*  $\alpha$ -SNAP alleles.

### **Wild-Type Soybean $\alpha$ -SNAPs Alleviate the Cytotoxicity and Secretion Defects of Resistance-Type $\alpha$ -SNAPs**

Due largely to two ancient genome polyploidization/duplication events (Schmutz et al., 2010), the Williams 82 soybean genome encodes five different  $\alpha$ -SNAPs: *Glyma.02G260400*, *Glyma.09G279400*, *Glyma.11g234500*, *Glyma.14G054900* and *Glyma.18g022500*. If resistance-type *Rhg1*  $\alpha$ -SNAPs interfere with NSF activities and vesicle trafficking, then the presence of these more canonical wild-type  $\alpha$ -SNAPs is likely to be crucial for the viability of soybeans carrying SCN resistance-conferring haplotypes of *Rhg1*. To determine if increased levels of wild-type  $\alpha$ -SNAPs could relieve the cytotoxicity of resistance-type *Rhg1*  $\alpha$ -SNAPs, we infiltrated a mixed culture of 3 parts  $\alpha$ -SNAP<sub>*Rhg1*</sub>LC to either 1 part WT- $\alpha$ -SNAP or 1 part empty vector. Even at this low ratio, co-expression of any of the highly similar Ch2, Ch11, Ch18 soybean  $\alpha$ -SNAPs - but not the divergent Ch9  $\alpha$ -SNAP - greatly diminished the cytotoxicity of resistance-type *Rhg1*  $\alpha$ -SNAP (Fig. 3A). Additionally, co-expression of 3 parts  $\alpha$ -SNAP<sub>*Rhg1*</sub>LC to 1 part WT- $\alpha$ -SNAP substantially rescued the exocytosis defect measured using sec-GFP secretion (Fig. 3B,C). Adjusting the ratio of co-infiltrated  $\alpha$ -SNAP<sub>*Rhg1*</sub>LC or WT  $\alpha$ -SNAP tipped the cell death vs. tolerance outcome in either direction (Fig. S5C). The above set of experiments

suggest that the cell death caused by resistance-type  $\alpha$ -SNAPs in these *N. benthamiana* assays is likely caused by overwhelming the endogenous  $\alpha$ -SNAPs with disruptive *Rhg1*  $\alpha$ -SNAPs.

Having shown that  $\alpha$ -SNAP<sub>*Rhg1*WT-L288A</sub> or  $\alpha$ -SNAP<sub>*Rhg1*WT-L288I</sub> expression alone was not cytotoxic in *N. benthamiana*, we tested if the penultimate leucine is required for wild-type  $\alpha$ -SNAP rescue of cell death from resistance-type  $\alpha$ -SNAPs. As in Fig 3A, we infiltrated a mixture of 3 parts  $\alpha$ -SNAP<sub>*Rhg1*LC</sub> to 1 part  $\alpha$ -SNAP<sub>*Rhg1*WT</sub>,  $\alpha$ -SNAP<sub>*Rhg1*WT-L288A</sub>,  $\alpha$ -SNAP<sub>*Rhg1*WT-L288I</sub> or empty vector control, respectively. As before, co-expressing  $\alpha$ -SNAP<sub>*Rhg1*WT</sub> diminished cytotoxicity. Co-expression of  $\alpha$ -SNAP<sub>*Rhg1*WT-L288A</sub> failed to relieve cell death (Fig. 3D).  $\alpha$ -SNAP<sub>*Rhg1*WT-L288I</sub> co-expression partially decreased cell death compared to empty vector co-infiltration, suggesting that NSF activation may be needed to prevent cell death and that a penultimate isoleucine, but not alanine, may confer partial  $\alpha$ -SNAP function.

### **During *Rhg1*-Mediated SCN Resistance, Resistance-type $\alpha$ -SNAPs Hyperaccumulate in SCN Feeding Sites and Reduce 20S Complexes**

We previously reported a positive correlation between the copy number of *Rhg1* repeats and higher levels of *Rhg1* transcripts (Cook et al., 2014). Because high ratios of resistance-type to wild-type  $\alpha$ -SNAPs disrupted trafficking and caused cell death in *N. benthamiana*, we examined if the balance of wild-type  $\alpha$ -SNAPs to resistance-type  $\alpha$ -SNAPs in soybean normally favors wild-type  $\alpha$ -SNAP activity, and is shifted specifically at the nematode feeding site during *Rhg1*-mediated resistance. Roots of non-transgenic soybean cultivar Fayette, which carries high-copy SCN resistance-type *Rhg1*, were inoculated with 200 juvenile SCN per root or mock inoculated. Four days later, SCN-infected root regions were isolated and pooled. Endogenous  $\alpha$ -SNAP<sub>*Rhg1*HC</sub>, WT- $\alpha$ -SNAP and NSF levels at the developing nematode-induced syncytium were

monitored using immunoblots. Substantial increases in  $\alpha$ -SNAP<sub>Rhg1</sub>HC and NSF protein abundance were detected in tissue enriched for SCN syncytia (feeding sites) (Fig. 4A). Moreover, the ratio of  $\alpha$ -SNAP<sub>Rhg1</sub>HC to WT- $\alpha$ -SNAP ratio in SCN infested vs. uninfected roots increased approximately 0.4, as revealed by densitometric analyses on (A) (Fig. 4B). SCN infestation of SCN-susceptible Williams82 roots did not reveal significant increases in WT- $\alpha$ -SNAP levels in SCN feeding sites as compared with uninfected controls (S10A).

Because pooling SCN-infected root regions includes considerable amounts of non-syncytial tissue, we utilized electron microscopy (EM) and immunogold labeling of  $\alpha$ -SNAP<sub>Rhg1</sub>HC to both pinpoint and more accurately assess  $\alpha$ -SNAP<sub>Rhg1</sub>HC protein elevation around the SCN feeding site in soybean cultivar Fayette roots. Immunogold labeling showed hyperaccumulation of the  $\alpha$ -SNAP<sub>Rhg1</sub>HC protein in syncytial cells but not in adjacent non-syncytial cells (Fig. 4C). Across three independent experiments, approximately 12-fold more immunogold particles were evident in syncytial cells relative to a similar two-dimensional area of adjacent cells (Fig. 4D, S10B). Anti- $\alpha$ -SNAP<sub>Rhg1</sub>HC immunogold particles were rare in non-infected samples. Fig. S10C,D shows images of SCN-infected and non-infected Fayette roots after contrasting, which clarifies cellular organelles but makes immunogold-labeled particles less obvious. To confirm antigen specificity of the anti- $\alpha$ -SNAP<sub>Rhg1</sub>HC antibody in EM/immunogold labeling usage, we conducted control experiments in which the antibody was pre-incubated with a 10-fold molar excess of purified  $\alpha$ -SNAP<sub>Rhg1</sub>HC protein before use on EM sections, and observed no staining in high-copy *Rhg1* roots (Fig. S10E). This indicates strong specificity of the antibody for the intended antigen in EM specimens. No immunogold labeling was observed when only the secondary antibody was used (Fig. S10F). Since  $\alpha$ -SNAP<sub>Rhg1</sub>HC hyperaccumulates in SCN feeding sites from Fayette, we examined if 20S complexes are

impacted during *Rhg1*-mediated resistance using density gradient centrifugation as in Fig. 1E. SCN-infested or mock inoculated root regions from Fayette were isolated and pooled four days after inoculation as in Fig 4A. Densitometric analysis of immunoblots indicates that NSF in 20S migrating fractions over total NSF abundance is specifically decreased in SCN-infested regions (Fig 4E,F). Together, these results show that *Rhg1*-mediated SCN resistance specifically triggers a shift to increased levels of  $\alpha$ -SNAP<sub>*Rhg1*</sub>HC in the syncytium, which likely impairs NSF function as evidenced by decreased 20S levels. We therefore propose that during SCN infection, resistant-type *Rhg1* haplotypes drive a localized hyperaccumulation of defective  $\alpha$ -SNAPs that inhibit NSF function and disrupt normal vesicular trafficking, interfering with pathogen co-option of cellular processes and reducing the viability of the syncytium SCN feeding site.

## 4.5 Discussion

The present study found that the agriculturally valuable *Rhg1* locus, which combats a highly damaging cyst nematode parasite, encodes disruptive  $\alpha$ -SNAP proteins that impair NSF function.  $\alpha$ -SNAP and NSF are core eukaryotic housekeeping genes that are central to SNARE recycling and vesicle trafficking. Our findings, that C-terminal polymorphisms in resistance-type soybean *Rhg1*  $\alpha$ -SNAPs reduce NSF interaction and 20S stability, disrupt vesicular trafficking and are cytotoxic, are consistent with animal studies on  $\alpha$ -SNAPs and NSF. In mice,  $\alpha$ -SNAP mutations such as the *hyh* allele are homozygous-lethal as are NSF-null *comatose* alleles in *Drosophila* (Sanyal and Krishnan, 2001; Chae et al., 2004). Artificial mutations at the penultimate C-terminal leucine of yeast or animal  $\alpha$ -SNAPs no longer stimulate NSF ATPase, impair SNARE recycling, block secretion and cause apoptosis in cell cultures (Barnard et al., 1996; Naydenov et al., 2012; Zick et al., 2015). What is unusual about soybean *Rhg1* is that sabotaging this core housekeeping function contributes to a beneficial trait - a trait that has been widely selected for by soybean breeders in recent decades to help control a disease that annually causes billions of dollars in lost food harvest worldwide.

Resistance through disruption of a core housekeeping process represents a departure from known mechanisms of plant disease resistance (Dangl et al., 2013; Niks et al., 2015). Ancient polyploidization in soybean (Schmutz et al., 2010) apparently allowed divergence of the *Rhg1*  $\alpha$ -SNAP gene to form an incompletely penetrant dominant-negative allele whose deleterious phenotype is dependent on the relative protein abundance of functional wild-type  $\alpha$ -SNAPs. We provide multiple lines of evidence demonstrating plant disease resistance that is promoted by a dysfunctional variant of an essential gene. Plant resistance to potyviruses is somewhat analogous in that it arises from mutations in the translation initiation factors eIF4G or eIF4E, which are

core housekeeping proteins (Kang et al., 2005). However, resistance-conferring eIF4 proteins provide a recessive resistance by precluding interactions with the potyvirus VPg, and otherwise appear to retain the normal activities of wild-type eIF4 proteins (Kang et al., 2005). Resistance to pathogens through compromises in essential gene function, partially analogous to *Rhg1*, have also been reported in humans. For example, resistance to malaria, and possibly typhoid fever, may be enhanced by specific mutations in hemoglobin or CFTR (cystic fibrosis transmembrane conductance receptor) respectively (Pier et al., 1998; Elguero et al., 2015). However, individuals homozygous for these alleles are afflicted with sickle cell anemia or cystic fibrosis. In the case of soybean *Rhg1*, the alleles that confer disease resistance are apparently tolerated because of polyploidization, with the effects of the resistance-conferring dysfunctional  $\alpha$ -SNAPs obscured in most tissues by wild-type  $\alpha$ -SNAP proteins produced by paralogous genes.

The hypothesis that *Rhg1* resistance-type  $\alpha$ -SNAPs interrupt NSF function and vesicle trafficking, yet are tolerated in high-yielding soybean varieties, is strongly supported by our findings that multiple wild-type  $\alpha$ -SNAPs mask the effects of resistance-type  $\alpha$ -SNAPs, even at low doses. It is further supported by the finding that the wild-type to resistance-type  $\alpha$ -SNAP ratio shifts during *Rhg1* mediated SCN resistance, with resistance-type  $\alpha$ -SNAPs hyperaccumulating in the SCN feeding site prior to its collapse. We do not yet know how the balance in SCN-infected tissues is tipped to an elevated presence of disruptive *Rhg1*  $\alpha$ -SNAP proteins. Elevated *Rhg1* gene expression in syncytia via transcription factor regulation is one obvious hypothesis, but other contributions may come from differential *Rhg1* locus methylation between haplotypes, dynamic infection-associated regulation of *Rhg1* locus methylation, miRNA-mediated transcriptional or post-transcriptional regulation, and syncytium-specific

genome endoreduplication (Song et al., 2011; de Almeida Engler and Gheysen, 2013; Yu et al., 2013; Cook et al., 2014)

A number of recent findings are at least partially consistent with the present finding of disruption of NSF functions and vesicular trafficking by resistance-type *Rhg1*  $\alpha$ -SNAPs. Microarray transcript abundance studies of laser-capture microdissected syncytium samples have indicated that *Rhg1*-mediated disease resistance is accompanied by a cellular stress profile that includes oxidative, cold, osmotic and unfolded protein stresses (Kandath et al., 2011). Additionally, the high metabolic demands and large scale membrane reorganizations necessary to form the syncytium (Niblack et al., 2006; Davis et al., 2008; Kyndt et al., 2013) may amplify cellular sensitivity to elevated levels of non-cooperative  $\alpha$ -SNAPs. A study of virulent SCN populations that had recently evolved to reproduce on soybeans carrying high-copy *Rhg1* haplotypes demonstrated allelic imbalance of a SNARE-like effector protein in the nematode (Bekal et al., 2015). Other recent reports suggest that a naturally occurring truncated soybean  $\alpha$ -SNAP enhances SCN resistance through increasing transcription of a Golgi localized SNARE, syntaxin-31 (Matsye et al., 2012; Pant et al., 2015). The amino acid sequence of that truncated  $\alpha$ -SNAP indicates it is encoded by *Glyma.11g234500* on chromosome 11, hence potential functional overlaps with *Rhg1*-encoded  $\alpha$ -SNAPs on chromosome 18 are unclear (Cook et al., 2014). The mechanisms by which the other *Rhg1*-encoded genes (Cook et al., 2012), *Glyma.18g022400* and *Glyma.18g022700*, contribute to *Rhg1*-mediated SCN resistance also remains unclear.

In the present study, high levels of *Rhg1* resistance-type  $\alpha$ -SNAPs triggered not only cytotoxicity, but also elevated levels of NSF protein. Regulation of NSF activity through post-translational phosphorylation has been reported, however, a cellular feedback mechanism that

adjusts NSF levels in response to  $\alpha$ -SNAP activity is apparently unreported and may represent a regulatory mechanism present in plants (Zhao et al., 2007). That  $\alpha$ -SNAP<sub>*Rhg1*WT</sub> L288A did not cause cell death but did elevate NSF levels suggests that this NSF feedback mechanism can, at least in the case of  $\alpha$ -SNAP<sub>*Rhg1*WT</sub> L288A, compensate for interference with NSF activity. The fact that stimulated increases in NSF abundance did not block the cytotoxicity of the resistance-type  $\alpha$ -SNAP<sub>*Rhg1*</sub> variants is consistent with the reduced NSF interaction and destabilization of 20S complexes observed for the resistance-type  $\alpha$ -SNAPs.

The finding that the more canonical soybean  $\alpha$ -SNAPs counteract the cytotoxicity of resistance-type  $\alpha$ -SNAPs suggests additional areas for study that may provide agriculturally useful findings. For example, it may be functionally relevant that low-copy *Rhg1* soybean haplotypes, which have been more difficult to couple with high grain yields, lack the single wild-type  $\alpha$ -SNAP-encoding *Rhg1* repeat present in high-copy *Rhg1* haplotypes (Cook et al., 2014). This may make lines carrying the low-copy haplotype more sensitive to negative effects of resistance-type  $\alpha$ -SNAPs. As another matter, the  $\alpha$ -SNAP<sub>*Rhg1*LC<sup>Splice</sup></sub> protein encoded by low-copy *Rhg1* was not observed to accumulate. Upregulation of the proportion of alternatively-spliced  $\alpha$ -SNAP<sub>*Rhg1*LC</sub> transcript could provide a bypass that reduces disruptive  $\alpha$ -SNAP production and promotes balance with regard to wild-type  $\alpha$ -SNAPs. Other areas for future work are suggested by the positive correlation between the strength of SCN resistance and copy number of high-copy ( $\alpha$ -SNAP<sub>*Rhg1*HC</sub>-encoding) *Rhg1* repeats (Cook et al., 2014; Lee et al., 2015). Improved SCN resistance may be obtained if haplotypes can be identified or generated that carry more *Rhg1* copies than the current mainstay ten-copy haplotype. With transgene or CRISPR/Cas9 technologies, it may be possible to more directly boost *Rhg1* effectiveness based on our findings regarding higher doses of resistance-type  $\alpha$ -SNAPs or substitutions at the penultimate  $\alpha$ -SNAP

residue. Extensive screens of soybean accessions carrying low-copy  $\alpha$ -SNAP<sub>Rhg1</sub>LC-encoding *Rhg1* haplotypes (Lee et al., 2015) have detected *Rhg1* copy numbers only at or below three. This suggests that with this more strongly cytotoxic  $\alpha$ -SNAP there may be a need to limit *Rhg1* copy number, to balance SCN resistance functions against the requirement for most tissues to contain a low relative dosage of dysfunctional  $\alpha$ -SNAPs in order to obtain healthy high-yielding soybean lines. But it may be possible to overcome this limitation by achieving more pronounced up-regulation of resistance-type  $\alpha$ -SNAP abundance at sites of SCN infection. More generally, the paradigm of disease resistance through high local expression of a toxic variant of a core housekeeping protein may be applicable to other host-pathogen interactions.

## 4.6 Materials & Methods

### Recombinant Proteins

ORFs for all *Rhg1*  $\alpha$ -SNAPs and soybean NSF, *Glyma.07G195900* and *Glyma.13G180100*, were cloned into the expression vector pRham N-His-SUMO Kan according to manufacturer's guidelines (Lucigen). Recombinant  $\alpha$ -SNAP<sub>*Rhg1*</sub>WT with the final 10 C-terminal residues truncated ( $\alpha$ -SNAP<sub>*Rhg1*</sub>WT(-10)) was generated from the pRham N-His-SUMO  $\alpha$ -SNAP<sub>*Rhg1*</sub>WT vector using the PIPE mutagenesis method to remove the final 10 codons. All expression constructs were chemically transformed into the expression strain "E. cloni 10G" (Lucigen), grown to OD<sub>600</sub> = 0.60 and then induced with 0.2% L-Rhamnose (Sigma) for ~8 hrs at 37° or overnight at 28°. Notably, recombinant production of the  $\alpha$ -SNAP<sub>*Rhg1*</sub>LC<sub>Splice</sub> protein required stringent expression conditions (induced at 18°C for ~10 hrs) compared with the other  $\alpha$ -SNAPs in order to recover any soluble protein. Purified recombinant mammalian His-NSF was a kind gift of Dr. Sebastian Bednarek (UW-Madison). Soluble, native recombinant His-SUMO- $\alpha$ -SNAPs or His-SUMO-NSF proteins were purified with Perfect Pro Ni-NTA resin (5Prime), with similar procedures described in (Hanson et al., 1997) though no subsequent gel filtration steps were performed. Following the elution of the His-SUMO-fusion proteins, over-night dialysis was performed at 4° in 20 mM Tris, pH 8.0, 150 mM NaCl, 10% glycerol and 1.5 mM TCEP. The His-SUMO affinity/solubility tags were cleaved from  $\alpha$ -SNAP or soybean NSF using 1-2 units of SUMO Express protease (Lucigen) and separated by rebinding of the tag with Ni-NTA resin and collecting the recombinant protein from the flow-through. Recombinant protein purity was assessed by Coomassie blue staining and quantified via a spectrophotometer.

### *In vitro* $\alpha$ -SNAP and NSF binding assays

*In vitro* NSF binding assays were performed essentially as outlined in (Barnard et al., 1996, 1997). Briefly, 20  $\mu$ g of each recombinant  $\alpha$ -SNAP protein was placed into a 1.5 mL polypropylene tube and incubated at room temperature for 20 mins. Unbound  $\alpha$ -SNAP was washed with SNAP wash buffer (25 mM Tris, pH. 7.4, 50 mM KCl, 1 mM DTT, 1 mg /mL BSA) and 20  $\mu$ g of recombinant NSF was added and incubated on ice for 10 mins. NSF was then removed and each sample was washed twice to remove unbound NSF. Samples were then boiled in 1X SDS loading buffer and separated onto an 8% SDS-PAGE and silver stained using the Proteosilver kit (Sigma-Aldrich) following manufacturer's guidelines. The amount of NSF bound to various  $\alpha$ -SNAPs was calculated by densitometric analysis with ImageJ.

### **Plasmid Constructs**

Transient overexpression of soybean  $\alpha$ -SNAPs or soybean NSF was performed using the previous described soybean ubiquitin promoter in the binary vector pSM101 (Cook et al., 2012) or with the 35S promoter from pGWB6 (Song et al., 2015). The soybean  $\alpha$ -SNAP ORFs for *Glyma.18g022500*, *Glyma.11g234500*, *Glyma.02G260400*, *Glyma.09G279400* and the soybean NSF ORFs for *Glyma.07G195900* and *Glyma.13G180100* were PCR amplified from Williams82, Fayette or Forrest cDNAs generated using the iScript cDNA Synthesis Kit (Bio-Rad) and KAPA HiFi polymerase (Kapa Biosystems). Each respective ORF was placed directly under the control of the soybean ubiquitin promoter in the vector pBlueScript using the polymerase incomplete primer extension (PIPE) method (Klock and Lesley, 2009) and sequence verified.  $\alpha$ -SNAP or NSF expression cassettes were digested with XbaI/PstI or SbfI/AvrII (New England Biolabs) and gel extracted using the Qiaquick gel extraction kit (Qiagen). Purified DNA fragments were then ligated into the binary vector pSM101 using T4 DNA ligase (New England Biolabs). Mutagenesis of  $\alpha$ -SNAPs to create penultimate residue substitutions, C-terminal

truncations or ATPase-null NSF constructs was also performed using PIPE-based mutagenesis with KAPA HiFi polymerase.

### **Transient *Agrobacterium* Expression in *Nicotiana benthamiana***

*Agrobacterium tumefaciens* strain GV3101 (pMP 90) containing specified expression constructs was syringe infiltrated at  $OD_{600} = 0.60$  (unless otherwise noted) into young leaves of ~4 week old *Nicotiana benthamiana* plants. All GV3101 cultures were grown overnight at 28°C in 25  $\mu\text{g mL}^{-1}$  kanamycin, rifampicin and induced for ~2.5 hrs in 10 mM MES pH 5.60, 10mM  $\text{MgCl}_2$  and 100  $\mu\text{M}$  acetosyringone prior to leaf infiltration. *N. benthamiana* plants were grown at 25°C with a photoperiod of 16 hrs light at 100  $\mu\text{Em}^2$  and 8 hrs dark. For  $\alpha\text{-SNAP}_{Rhg1}\text{LC}$  complementation with WT  $\alpha\text{-SNAP}$  co-infiltration, 3 volumes of  $\alpha\text{-SNAP}_{Rhg1}\text{LC}$   $OD_{600} = 0.60$  were well-mixed with 1 volume of the specified WT  $\alpha\text{-SNAP}$  at  $OD_{600} = 0.60$ , or empty vector  $OD_{600} = 0.60$  immediately prior to co-infiltration. For sec-GFP co-expression experiments, sec-GFP was co-infiltrated at  $OD_{600} = 0.015$  with a specified *Rhg1*  $\alpha\text{-SNAP}$  at  $OD_{600} = 0.60$  or empty vector at  $OD_{600} = 0.60$  (Batoko et al., 2000) . For co-IP analysis, soybean  $\text{NSF}_{\text{Ch7}}\text{-HA}$  cultures were mixed with either  $\text{GFP-}\alpha\text{-SNAP}_{Rhg1}\text{HC}$  or  $\text{GFP-}\alpha\text{-SNAP}_{Rhg1}\text{WT}$  and co-infiltrated at  $OD_{600} = 0.40$  for each construct.

### **Co-immunoprecipitation**

cDNAs of  $\text{NSF}_{\text{Ch7}}$  and  $\alpha\text{-SNAP}_{Rhg1}\text{HC}$  or  $\alpha\text{-SNAP}_{Rhg1}\text{WT}$  were cloned into the HA-tagged pSM101 (soybean ubiquitin promoter) and GFP-tagged pGWB6 (35S promoter) vectors, respectively, and transformed into *A. tumefaciens* GV3101 (pMP90). Leaves of four week-old *N. benthamiana* plants were agroinfiltrated at  $OD_{600} 0.4$  and leaf tissues were harvested three days later. Total proteins were extracted in lysis buffer (50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 5 mM EDTA, 0.2% Triton X-100, 10% glycerol, and plant protease inhibitor cocktail (Sigma) at

1:100). Immunoprecipitation was carried out as described in (Song et al., 2015) with anti-GFP (Abcam) antibody at 4°C overnight followed by incubation with protein A beads (Thermo Scientific) for 1-2 hrs. The beads were washed three times with extraction buffer without protease inhibitors. The precipitated proteins were eluted with 1X SDS loading buffer, subjected to SDS-PAGE and immunoblotted with anti-HA (Roche) and anti-GFP (Clontech) antibodies, and detected using Supersignal West Pico or Dura chemiluminescent substrates (Thermo Scientific).

### **Glycerol Gradient Ultracentrifugation and Fractionation**

Glycerol gradient ultracentrifugation was performed similarly to (Bassham and Raikhel, 1999; Rancour et al., 2002). 20S complex abundance was quantified as the amount of NSF in ~20S migrating complexes over the total amount of NSF present and was calculated by densitometric analysis of NSF band intensity using ImageJ (Schindelin et al., 2012). Gradients of 40%-17.5% glycerol (V/V) were layered into 13 x 51mm Ultra-Clear tubes (Beckman Coulter) and allowed to settle at 4°C 1 hr prior to use. Transgenic *N. benthamiana* leaves were harvested at 3 days post infiltration. Leaf lysates were prepared as outlined for *Arabidopsis* roots in (Bassham and Raikhel, 1999) and membrane pellets were detergent-solubilized in a gradient buffer non-permissive to ATP hydrolysis (20 mM HEPES, pH 7.50, 50 mM KCl, 2 mM EDTA, 2 mM DTT, 1 mM ATP and 1% Triton X-100). Equal amounts of solubilized membrane proteins, as determined by Bradford assay, were then layered onto the gradients and separated by centrifugation at 125,000 g for 18 hrs in a MLS-50 swinging bucket rotor (Beckman Coulter). 400 uL fractions were collected from the top, except for the final 100 uL fractions which included pellet material and were excluded from final analyses. Fraction sedimentation was

monitored by parallel co-fractionation of protein standards of known sedimentation: BSA (4.4S), alcohol dehydrogenase (7.6S), catalase (11.3S) and thyroglobulin (19.4S) (S4).

### **Antibody Production**

Affinity purified, polyclonal antibodies were raised against synthetic peptide sequences matching the final six or seven C-terminal  $\alpha$ -SNAP residues: “EEDDLT” , “EQHEAIT” or “EYEVIT”, for wild-type, high-, or low-copy  $\alpha$ -SNAPs, respectively. For soybean NSF, a synthetic peptide “ETEKNVRDLFADAEQDQRTRGDESD”, matching residues 300-324 was used. Resistance-type  $\alpha$ -SNAP peptides and antibodies were produced by New England Peptides, while NSF and wild-type  $\alpha$ -SNAP antibodies were produced by Pacific Immunology. Antibody specificity was validated through immunoblots using various recombinantly produced  $\alpha$ -SNAP and NSF proteins and also root lysates of high or low copy *Rhg1*-containing lines, transgenic *N. benthamiana* leaves expressing various  $\alpha$ -SNAPs or Williams82 (single copy) hairy roots expressing various  $\alpha$ -SNAPs (S5A-D).

### **Immunoblots**

Soybean roots or *N. benthamiana* leaf tissue was flash frozen in liquid nitrogen and homogenized in 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 5 mM EDTA, 0.2% Triton X-100, 10% glycerol, and protease inhibitor cocktail in a Power Lyzer 24 (Mo Bio) for 3 cycles at 15 seconds each, with flash freezing in between cycles. Immunoblots for either *Rhg1*  $\alpha$ -SNAP were incubated overnight at 4°C in 5% non-fat dry milk TBS-T (50 mM Tris, 150 mM NaCl, 0.05% Tween-20) at 1:1,000. NSF immunoblots were performed similarly, except incubations were 1 hr at room temperature. Secondary horseradish peroxidase conjugated goat-anti-rabbit IgG and incubated at 1:10,000 for 1 hr at room temperature on a platform shaker. Chemiluminescent

detection was performed with SuperSignal West Chemiluminescent Substrate Pico or Dura (Thermo Scientific) and developed using a ChemiDoc MP chemiluminescent imager (Bio-Rad).

### **Transgenic Soybean Hairy Root Production**

Transgenic soybean hairy roots were generated by transformation of soybean with *Agrobacterium rhizogenes* strain Arqua1 as described in (Cook et al., 2012).

### **Confocal Microscopy**

Live-cell imaging experiments were performed using an inverted laser scanning confocal microscope (Elyra LSM 780; Carl Zeiss) with a 20X or 40X water immersion objective. Transformed leaves were analyzed 72 hrs after infiltration. The excitation wavelength for GFP was 488 nm and the emitted fluorescence was collected with either 510–525 nm emission filter. Individual experiments for sec-GFP fluorescence were performed by single-imaging frame collection using identical laser output levels and imaging conditions on cells expressing sec-GFP or coexpressed with empty vector or indicated soybean SNAPS. Images were captured using a standardized scan area of 442.2 x 442.2  $\mu\text{m}$  (pixel size 0.87  $\mu\text{m}$ ), with a frame size of 512 x 512 and a scan time of 968.14 msec. The 488 nM laser intensity was set at 2.5 with a master gain setting of 725 and a pinhole of 32.3 (0.84 Airy Units). At least 25 images were taken for each expression construct. Sec-GFP fluorescence was quantified using Fiji software (Schindelin et al., 2012). GFP fluorescence intensity was calculated by highlighting each image with the *Rectangular Selection* tool and analyzing for mean pixel intensity of the total epidermal cell surface area ( $\mu\text{m}^2$ ). Syp61-mCherry imaging was performed similarly, except image collection was performed with a 40X wet-mount on mesophyll cells. The excitation wavelength for mCherry was 561 nm. Four separate plants from 3 independent experiments were used and >50 images of each treatment were collected.

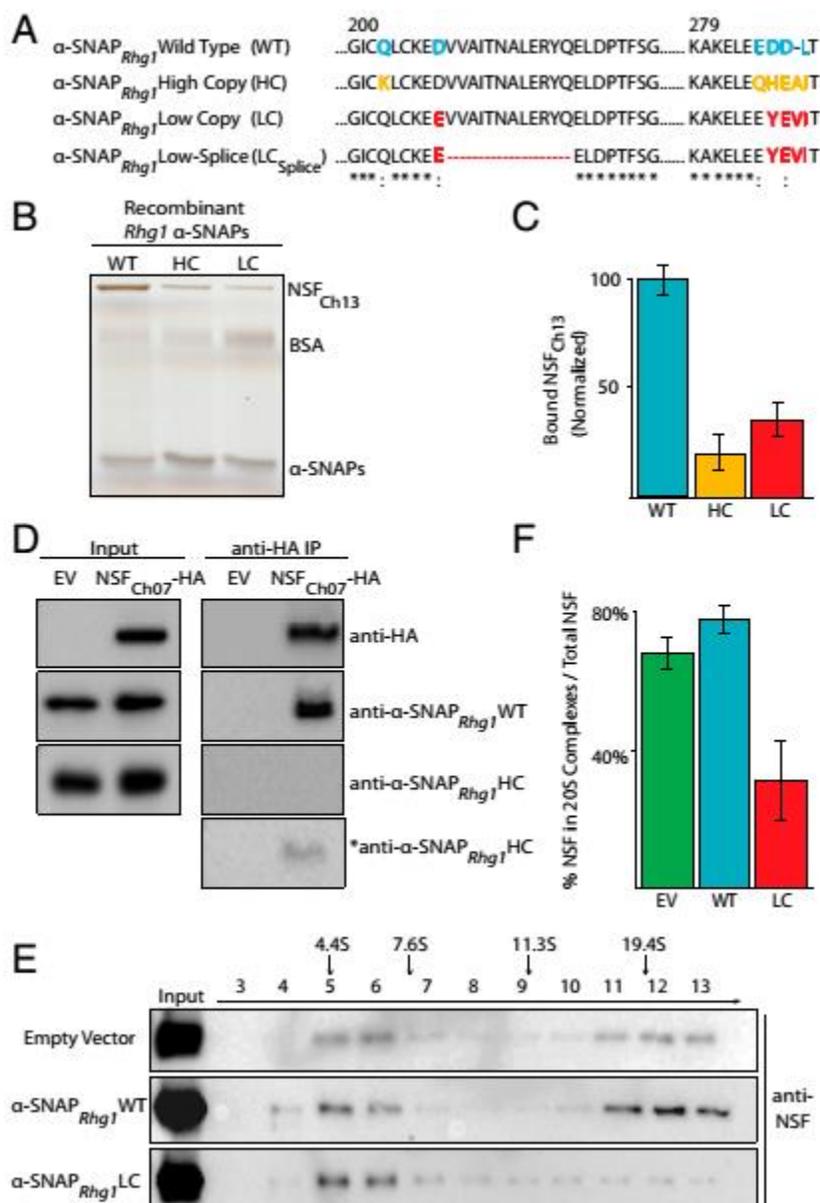
## **Electron Microscopy**

Syncytia from soybean roots (Fayette) inoculated with 200 juvenile stage 2 (J2) SCN (Race 0) were hand sectioned with a razor at 4 dpi. Root sections were fixed in 0.1% glutaraldehyde and 4% paraformaldehyde in 0.1 M sodium phosphate buffer (PB), pH 7.4 overnight (under vacuum for the first hr). Samples were washed four times with 0.1 M PB, dehydrated in ethanol and embedded in LR White. Ultrathin sections (~90 nm) were taken using an ultramicrotome (Leica UC-6) and mounted on nickel slot grids. For the immunogold labeling procedure, grids were incubated on drops of 50 mM glycine/PBS for 15 min followed by drops of prepared blocking buffer (Aurion) for 30 min and then equilibrated in 0.1% BSA-C/PBS (incubation buffer) (Aurion). Next, grids were incubated with the indicated antibodies diluted 1:200 (in incubation buffer) overnight at 4°C, washed five times in incubation buffer, and incubated for 2 hrs with goat anti-rabbit antibody conjugated to 15 nm gold (Aurion) diluted 1:25 in incubation buffer. After six washes in incubation buffer, two 5-min washes in PBS, the grids were fixed for 5 min in 2.0% glutaraldehyde in 0.1 M phosphate buffer, followed by two 5-min washes in 0.1 M phosphate buffer, and five 2-min washes in water. Finally, the grids were contrasted with 2% aqueous uranyl acetate and Reynolds lead citrate. Images were collected with a MegaView III digital camera on a Phillips CM120 transmission electron microscope.

#### **4.7 Acknowledgements**

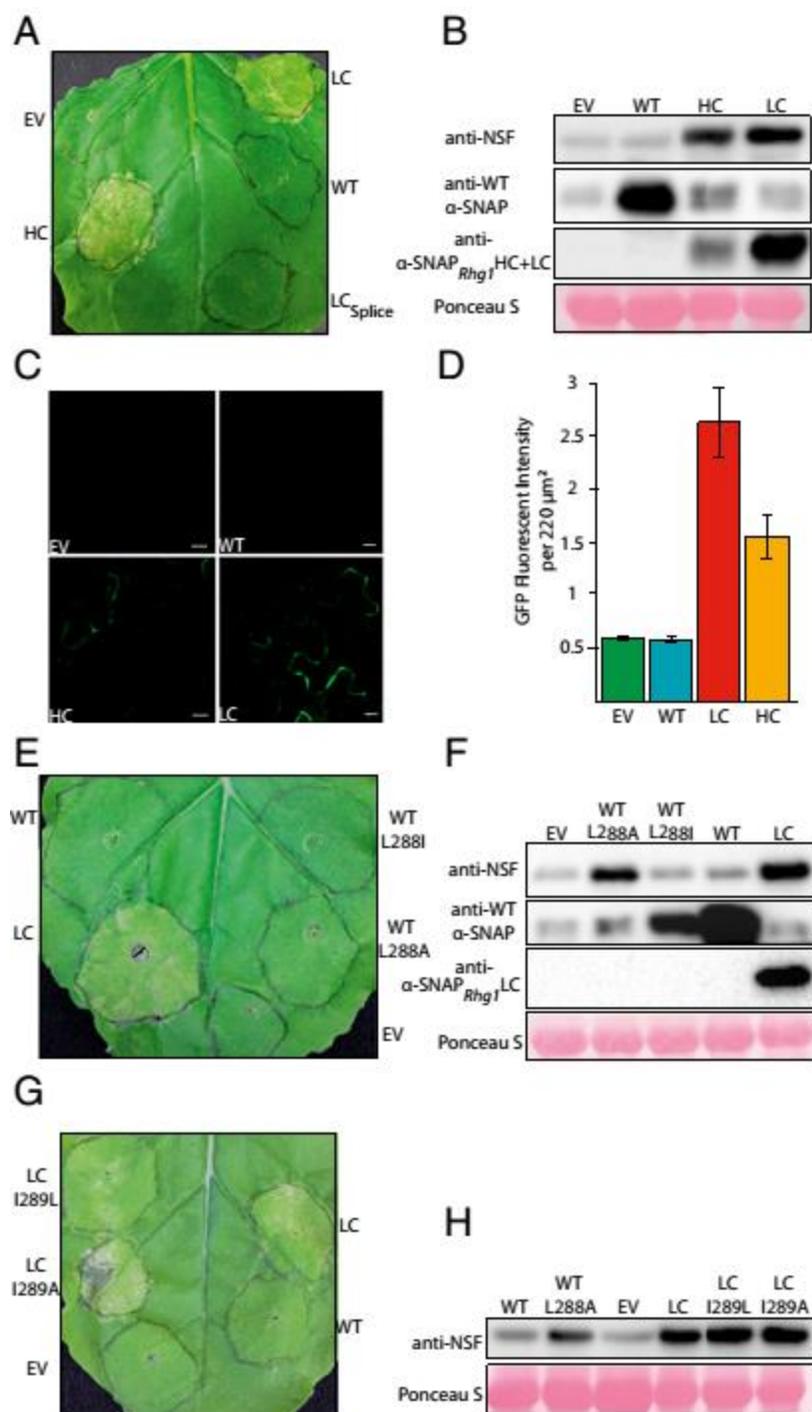
We thank Dr. Sebastian Bednarek for sharing recombinant mammalian NSF<sub>CHO</sub> proteins and substantial guidance, as well as Dr. Declan James and Dr. Marisa Otegui for suggestions. We also thank Dr. Roger Innes (Indiana University) for sharing the Syp61-mCherry construct and Dr. Hugo Zheng (McGill University) for providing the sec-GFP construct. This work was funded primarily by USDA-NIFA-AFRI award 2014-67013-21775 to A.F.B., and also by the United Soybean Board. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. (DGE-1256259) to A.M.B.

## 4.8 Figures



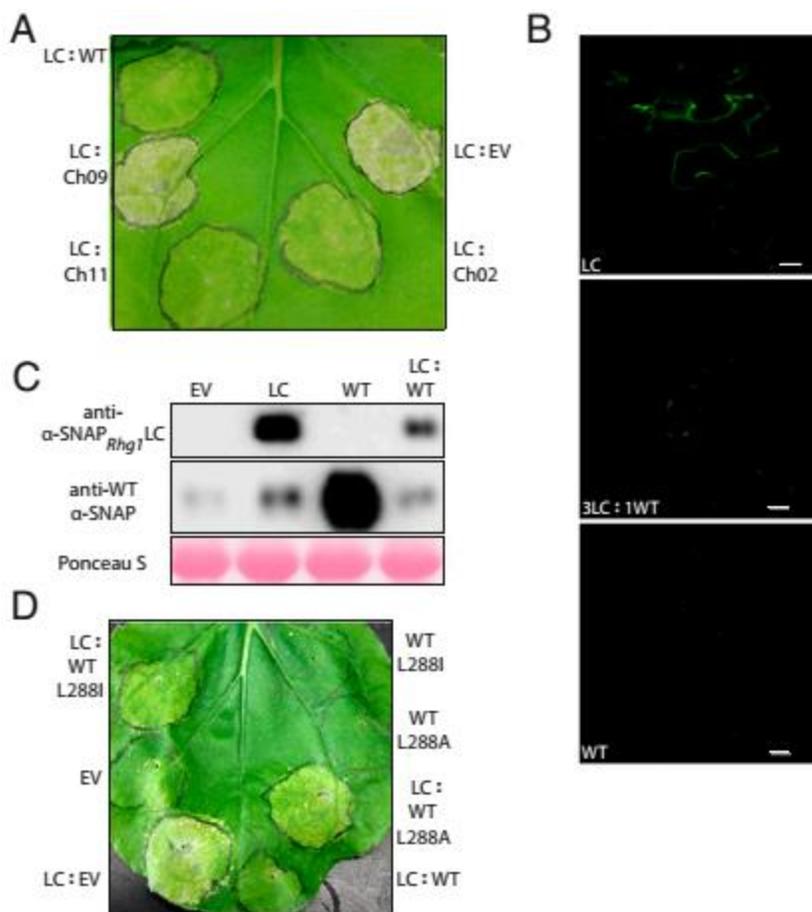
**Fig. 1.** *Rhg1* resistance-type  $\alpha$ -SNAPs are deficient in NSF interactions and destabilize 20S complexes (A) Alignment of *Rhg1* single-copy (wild-type, SCN susceptible), low-copy (SCN resistant) and high-copy (SCN resistant)  $\alpha$ -SNAPs (Cook et al., 2014), showing resistance-type amino acid polymorphisms and an alternate splice form of the low copy  $\alpha$ -SNAP. (B) Silver stained SDS-PAGE of recombinant soybean NSF<sub>Ch13</sub> bound *in vitro* by recombinant wild-type

(WT), low-copy (LC), or high-copy (HC) *Rhg1*  $\alpha$ -SNAP proteins. (C) Densitometric quantification of NSF<sub>Ch13</sub> bound by *Rhg1*  $\alpha$ -SNAPs as in (B); data from three independent NSF<sub>Ch13</sub> experiments; error bars show SEM. (D) NSF co-immunoprecipitation upon anti-GFP immunoprecipitation (IP) of GFP- $\alpha$ -SNAP<sub>*Rhg1*WT</sub> or GFP- $\alpha$ -SNAP<sub>*Rhg1*HC</sub> co-expressed with soybean NSF<sub>Ch07</sub>-HA in *N. benthamiana* leaves. Input: total protein samples, prior to immunoprecipitation. (E) Immunoblot of density gradient fractions to detect presence of NSF in 20S complexes. Total solubilized membrane proteins were loaded from *N. benthamiana* leaves expressing either  $\alpha$ -SNAP<sub>*Rhg1*LC</sub>,  $\alpha$ -SNAP<sub>*Rhg1*WT</sub> or empty vector, and anti-NSF antibody was used to detect endogenous *N. benthamiana* NSF after SDS-PAGE immunoblot of the resulting fractions. (F) Quantification of NSF present in 20S complexes; densitometric data from four independent experiments, calculated as the combined density of NSF signal in ~20S migrating fractions (fractions 11-13) over the total NSF density (fractions 3-13). Error bars show SEM.

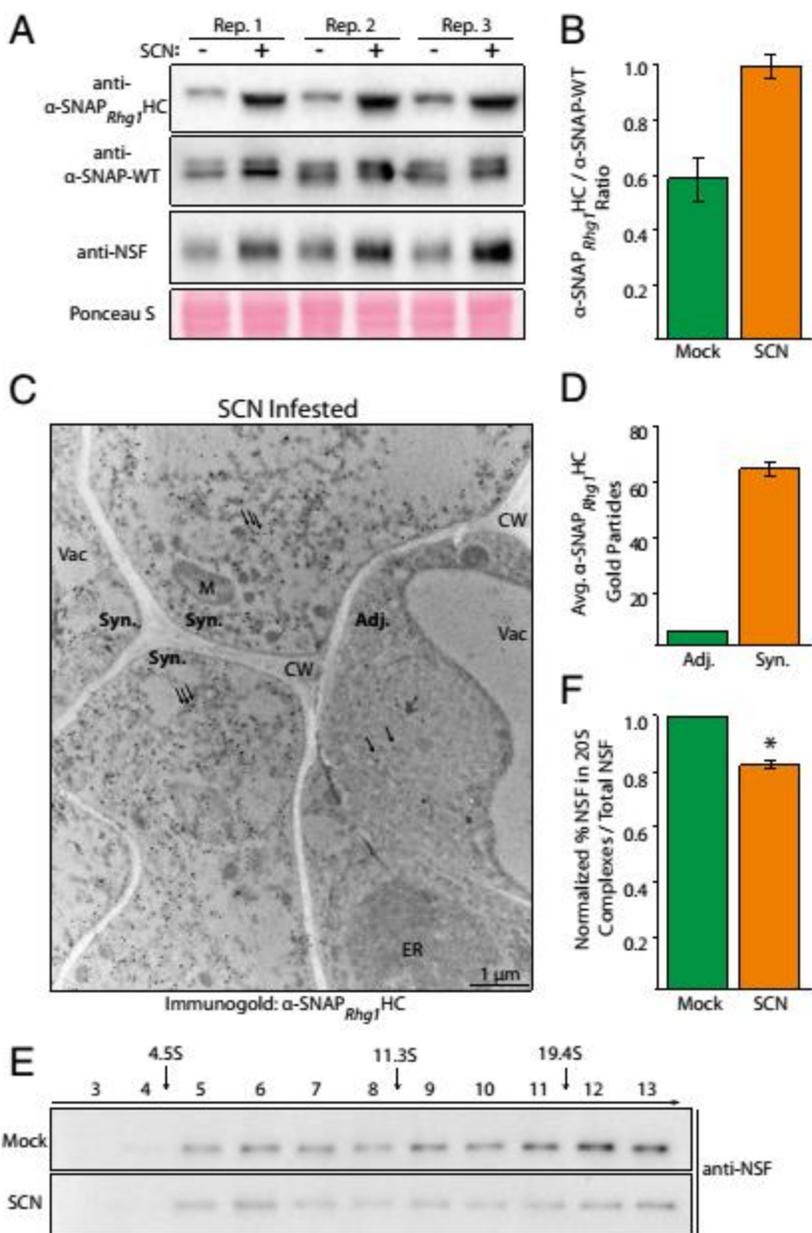


**Fig. 2.** *Rhg1* resistance-type  $\alpha$ -SNAP expression disrupts secretory trafficking, triggers NSF hyperaccumulation and eventually causes cell death in *N. benthamiana*. (A) *N. benthamiana* leaf expressing *Rhg1*  $\alpha$ -SNAPs with no epitope tag, or an empty vector control, six days after

agroinfiltration. EV: empty vector; WT:  $\alpha$ -SNAP<sub>Rhg1</sub>WT; LC:  $\alpha$ -SNAP<sub>Rhg1</sub>LC; HC:  $\alpha$ -SNAP<sub>Rhg1</sub>HC; LC<sub>Splice</sub>:  $\alpha$ -SNAP<sub>Rhg1</sub>LC<sub>Splice</sub>. (B) Immunoblot of endogenous *N. benthamiana* NSF abundance in leaves expressing the indicated  $\alpha$ -SNAP<sub>Rhg1</sub> constructs from (A) or empty vector control. Same samples probed with anti- $\alpha$ -SNAP<sub>Rhg1</sub> antibodies raised against peptides from the indicated source. Leaf tissue harvested three days after agroinfiltration; Ponceau S stain for similar loading of total protein. (C) Confocal images of *N. benthamiana* epidermal cells co-expressing sec-GFP and *Rhg1*-encoded  $\alpha$ -SNAPs denoted as in (A), or empty vector (EV). Sec-GFP assay detects GFP signal if there is failed secretion (retention in ER/Golgi). Images for three days after agroinfiltration; scale bars indicate 20  $\mu$ m. (D) Quantification of sec-GFP fluorescence with the respective *Rhg1*-encoded  $\alpha$ -SNAPs as shown in (C) using ImageJ; n=25 for each construct; error bars are SEM. (E,G) *N. benthamiana* leaves five days after agroinfiltration to express the indicated *Rhg1*  $\alpha$ -SNAPs with no epitope tag, or *Rhg1*  $\alpha$ -SNAPs mutagenized to carry different residues at the penultimate amino acid (no epitope tag), or an empty vector control. (F,H) Endogenous *N. benthamiana* NSF abundance at three days as in (B), upon expression of the indicated  $\alpha$ -SNAP<sub>Rhg1</sub> constructs from (E) or (G) respectively, or empty vector control.



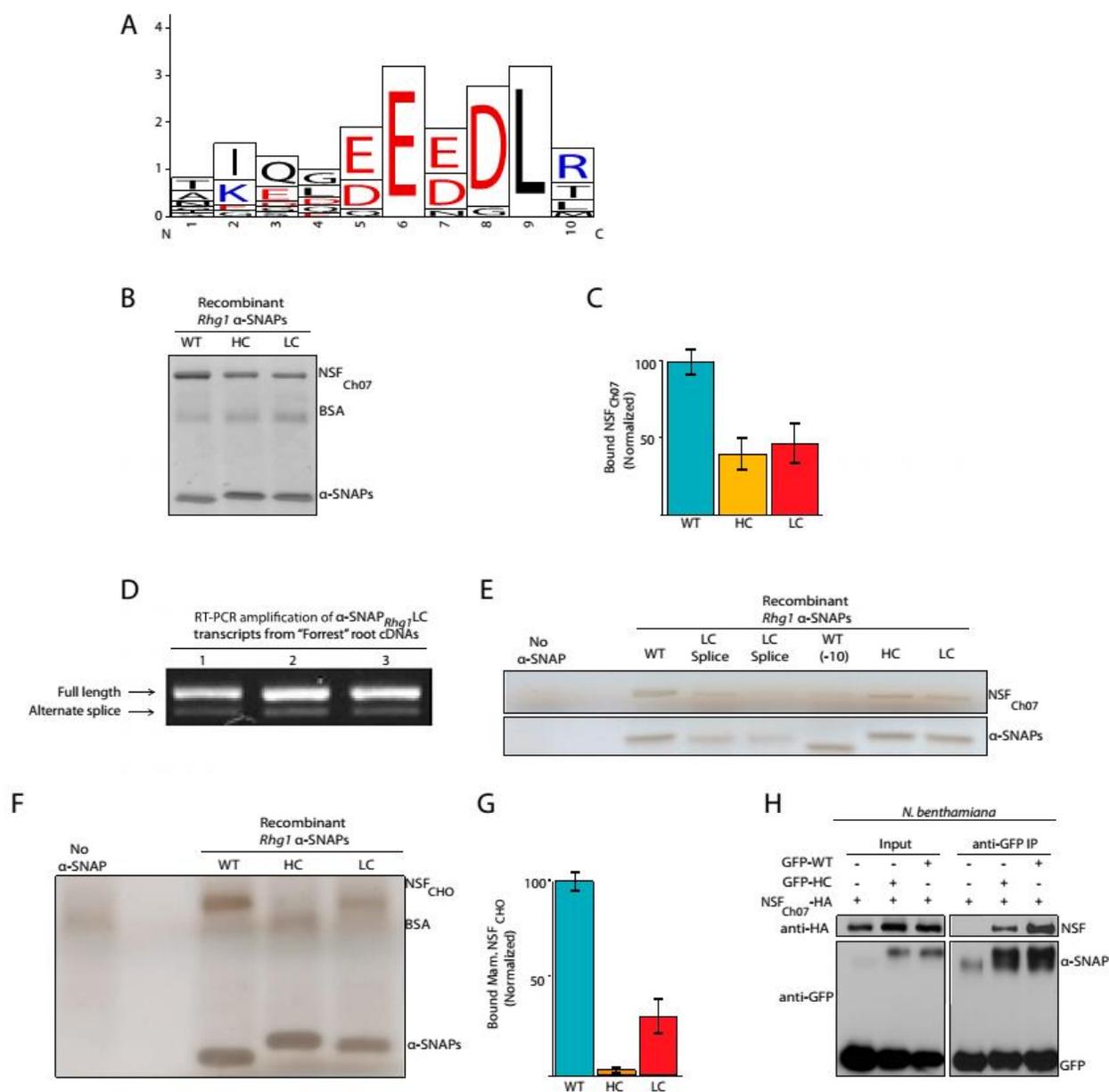
**Fig. 3.** Coexpression of wild-type soybean  $\alpha$ -SNAPs with *Rhg1* resistance-type  $\alpha$ -SNAPs alleviates cell death symptoms and secretion defects; penultimate leucine required. (A) *N. benthamiana* leaves six days after agroinfiltration with a 3:1 *Agrobacterium* culture mixture (three parts  $\alpha$ -SNAP<sub>*Rhg1*</sub>LC to one part wild-type soybean  $\alpha$ -SNAP or empty vector control). The soybean wild-type  $\alpha$ -SNAPs are: WT,  $\alpha$ -SNAP<sub>*Rhg1*</sub>WT = *Glyma.18g022500*; Ch<sub>02</sub>: *Glyma.02G260400*; Ch<sub>11</sub>: *Glyma.11g234500*; Ch<sub>09</sub>: *Glyma.09G279400*. (B) Confocal imaging of sec-GFP assays as in Fig 2C, but including leaves treated with 3:1 *Agrobacterium* culture mixture as in (A). (C) Immunoblot of leaf samples taken three days after agroinfiltration as in (A); LC : WT constructs infiltrated at 3:1 ratio. (D) Similar to (A), with 3:1 culture mixture of  $\alpha$ -SNAP<sub>*Rhg1*</sub>LC to either  $\alpha$ -SNAP<sub>*Rhg1*</sub>WT, or  $\alpha$ -SNAP<sub>*Rhg1*</sub>WT with A or I penultimate residue substitutions, or empty vector control.



**Fig. 4.**  $\alpha$ -SNAP<sub>Rhg1</sub>HC hyperaccumulates relative to WT  $\alpha$ -SNAPs at SCN infection sites in high-copy *Rhg1* soybean accession Fayette. (A) Immunoblot of tissue samples from SCN-infested root regions, harvested four days after SCN infection. Blots probed with the indicated antibodies; quantitative comparisons are valid within rows but not within columns. (B) Densitometric ratio of  $\alpha$ -SNAP<sub>Rhg1</sub>HC to WT  $\alpha$ -SNAPs calculated from band intensities in (A).

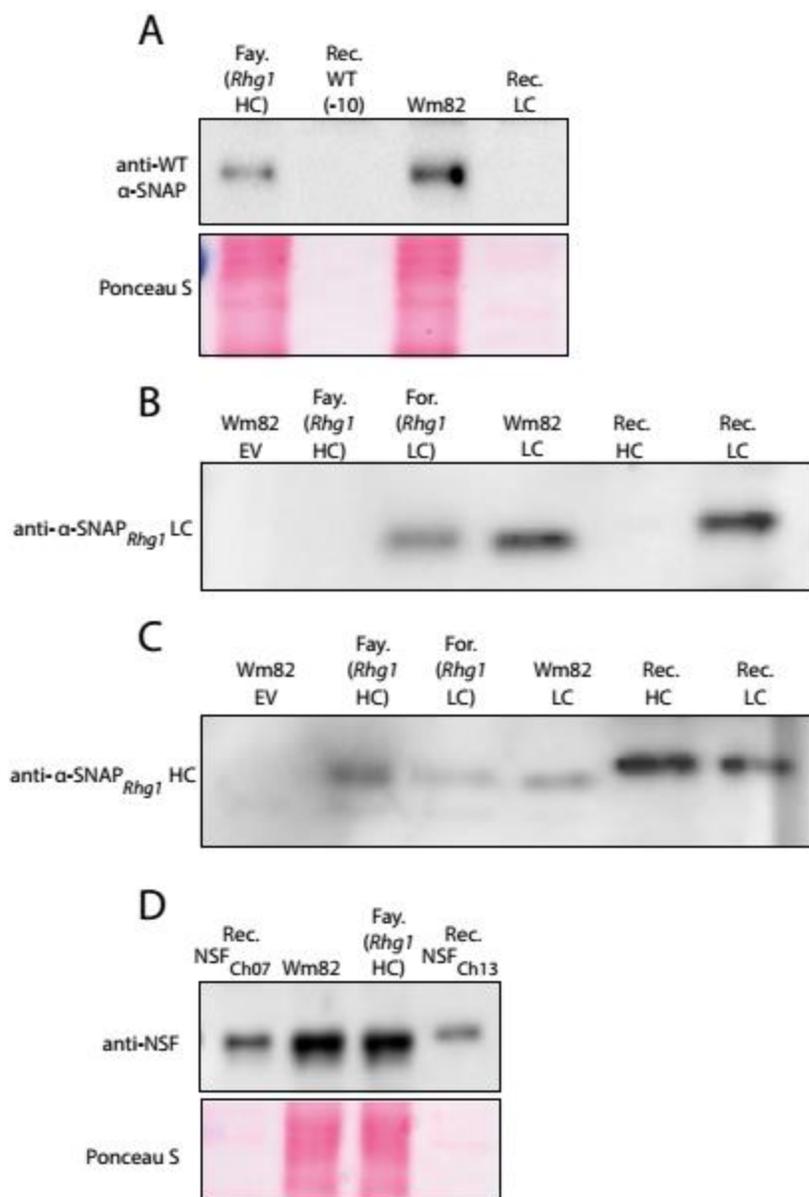
Error bars represent SEM. (C) Brightness adjusted electron micrograph showing immunogold-labeled  $\alpha$ -SNAP<sub>Rhg1</sub>HC in syncytia cells (Syn.) and adjacent cells (Adj.) four days after SCN infection of high copy *Rhg1* soybean accession Fayette. Arrows highlight eight of the approximately 400 immunogold particles in this image. M: mitochondrion; Vac: vacuole; ER: endoplasmic reticulum; CW: cell wall. (D) Average  $\alpha$ -SNAP<sub>Rhg1</sub>HC immunogold particle counts in syncytia vs. adjacent cells from 30 images across 3 independent experiments. See Fig. S10 for raw immunogold particle counts, additional images and antibody specificity controls. (E) anti-NSF immunoblot of density gradient fractions to detect 20S complexes from SCN-infested Fayette root regions, harvested four days after SCN infection. (F) Densitometric analysis of NSF from 20S migrating fractions (fractions 11-13) over total NSF (fractions 3-13) in SCN-infested root regions. Data from three independent experiments normalized to 20S NSF abundance from uninfected roots regions,  $p = 0.0225$ , paired T-test.

## 4.9 Supplemental Figures



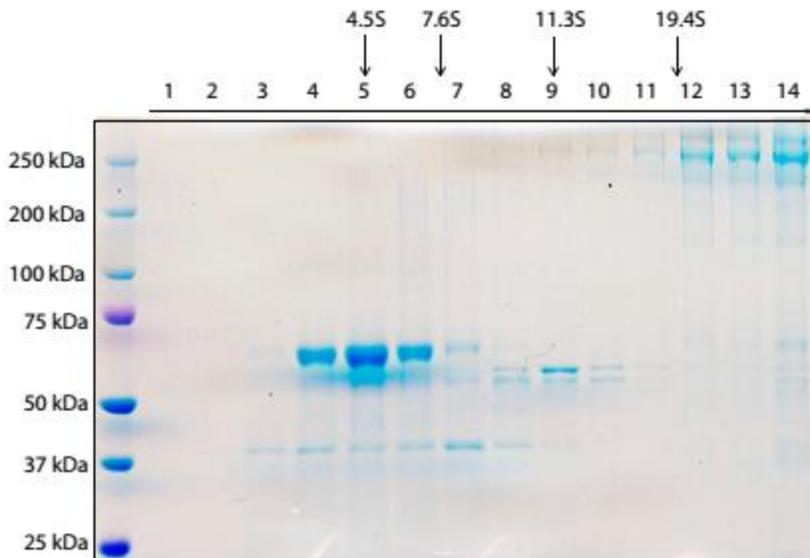
**Fig. S1.**  $\alpha$ -SNAP extreme C terminus is highly conserved among eukaryotes and critical for NSF binding. (A) Logo showing conservation of the final 10 C-terminal  $\alpha$ -SNAP residues from model organisms across diverse phyla, similar to (Cook et al., 2014)  $\alpha$ -SNAP C-terminal consensus was generated from the following species: *Chlamydomonas reinhardtii*, *Saccharomyces cerevisiae*,

*Physcomitrella patens*, *Arabidopsis thaliana*, *Medicago truncatula*, *Nicotiana tabaccum*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Danio rerio*, *Xenopus laevis*, *Gallus gallus*, *Rattus norvegicus*, and *Homo sapiens*. The conservation logo was generated using WebLogo (Crooks et al., 2004). (B) Silver-stained SDS/PAGE of recombinant soybean NSF<sub>Ch07</sub> bound in vitro by recombinant wild-type (WT), low-copy (LC), or high-copy (HC) *Rhg1*  $\alpha$ -SNAP proteins. (C) Densitometric quantification of NSF<sub>Ch07</sub> bound by *Rhg1*  $\alpha$ -SNAPs as in Fig. 1C; data are from four independent NSF<sub>Ch07</sub> experiments. Error bars show SEM. (D) Agarose gel showing RT-PCR product generated due to the presence of both full-length transcript and the alternate splice product. RT-PCR was performed on low-copy *Rhg1* line Forrest cDNA with a primer directly upstream of the splice site and at a sequence unique to the low-copy *Rhg1*  $\alpha$ -SNAP C terminus. Alternate splicing represents roughly 20% of total low-copy  $\alpha$ -SNAP transcripts. (E) Silver-stained SDS/PAGE of recombinant soybean NSF<sub>Ch07</sub> bound to recombinant *Rhg1*  $\alpha$ -SNAPs, including the alternately spliced low-copy  $\alpha$ -SNAP protein (LC<sub>Splice</sub>) or a 10-residue C-terminal truncation of  $\alpha$ -SNAP<sub>*Rhg1*WT</sub> [WT(-10)]. (F) As in B, but with a *Rhg1*-encoded  $\alpha$ -SNAP binding assay with recombinant Chinese hamster ovary NSF (NSF<sub>CHO</sub>). (G) Densitometric analysis of in vitro NSF<sub>CHO</sub> binding from four independent experiments. Error bars show SEM. (H) NSF coimmunoprecipitation upon anti-GFP immunoprecipitation (IP) of GFP- $\alpha$ -SNAP<sub>*Rhg1*WT</sub> or GFP- $\alpha$ -SNAP<sub>*Rhg1*HC</sub> co-expressed with soybean NSF<sub>Ch07</sub>-HA in *N. benthamiana* leaves. Input: total protein samples before immunoprecipitation.

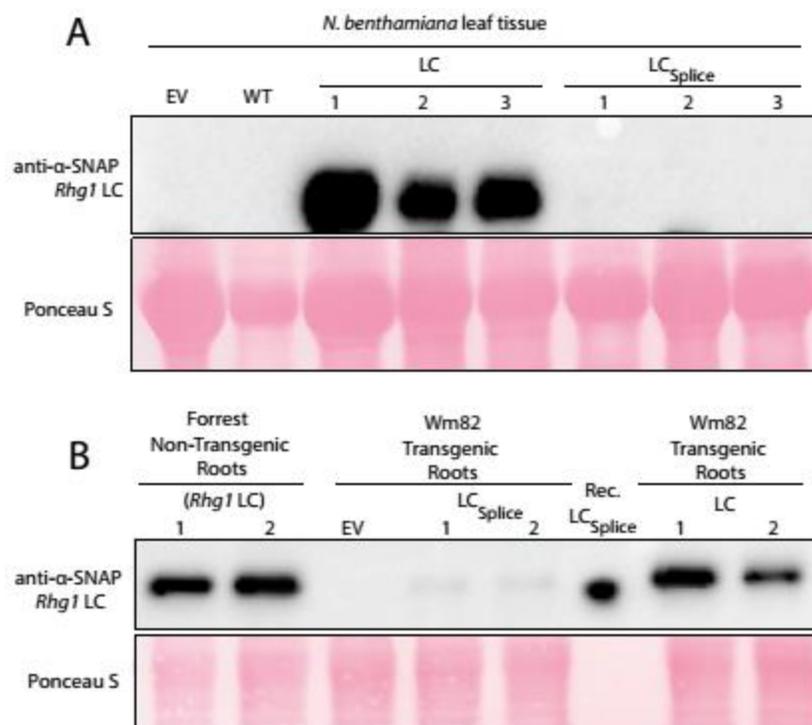


**Fig. S2.** Confirming the specificity of custom-generated  $\alpha$ -SNAP and NSF antibodies. (A) Immunoblot test of anti- $\alpha$ -SNAP WT on root lysates from Fayette(Fay.) or Williams 82 (Wm82), recombinant WT  $\alpha$ -SNAP truncated at the final 10 C-terminal residues and thereby lacking the epitope region [Rec. WT(-10)], or recombinant  $\alpha$ -SNAP<sub>*Rhg1*</sub>LC protein (Rec. LC). Note:  $\alpha$ -SNAP WT antibody was raised to the highly conserved  $\alpha$ -SNAP C terminus and is thus

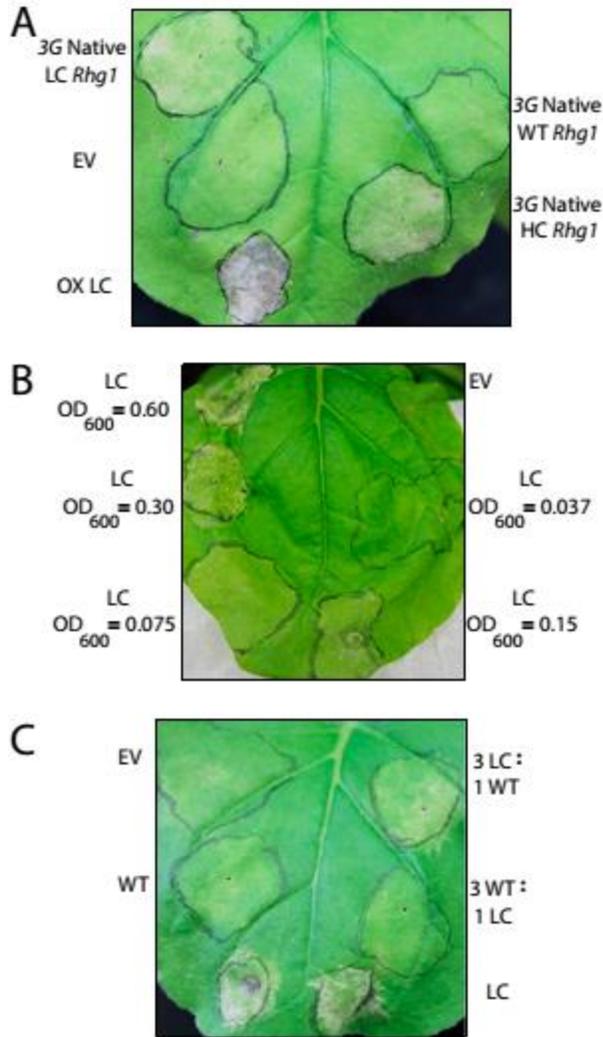
cross-reactive with most WT  $\alpha$ -SNAPs. (B) Immunoblot test of anti- $\alpha$ -SNAP<sub>Rhg1LC</sub> (low-copy) on root lysates from Fayette (endogenous high-copy *Rhg1*), Forrest (endogenous low copy *Rhg1*), or transgenic Williams 82 (single-copy *Rhg1*) roots expressing  $\alpha$ -SNAP<sub>Rhg1LC</sub> or an empty vector control (EV), or purified recombinant  $\alpha$ -SNAP<sub>Rhg1LC</sub> or recombinant  $\alpha$ -SNAP<sub>Rhg1HC</sub> protein. (C) Similar to B, but an immunoblot test of anti- $\alpha$ -SNAP<sub>Rhg1HC</sub> (high-copy). Note:  $\alpha$ -SNAP<sub>Rhg1HC</sub> antibody is cross reactive with  $\alpha$ -SNAP<sub>Rhg1LC</sub> but not with WT  $\alpha$ -SNAPs. (D) Immunoblot test of anti-NSF on recombinant NSF<sub>Ch07</sub> or NSF<sub>Ch13</sub> or root lysates from Fayette or Williams 82. As expected, the anti-soybean NSF antibody is also cross-reactive with the *N. benthamiana* NSF protein (e.g., Fig. 2).



**Fig. S3.** Density gradient fractionation of protein standards of known sedimentation, performed in the same run as one of the fractionations that detected the presence of NSF in 20S complexes (e.g., Fig. 1G). Sedimentation was performed similar to (Bassham and Raikhel, 1999) and (Rancour et al., 2002). Protein standards were detected by SDS/PAGE and Coomassie blue stain. Protein standards used were thyroglobulin (19.4S), ~250-kDa dimer; catalase (11.3S), ~60-kDa tetramer; yeast alcohol dehydrogenase (7.6S), ~37-kDa tetramer; and BSA (4.5S), ~65 kDa.

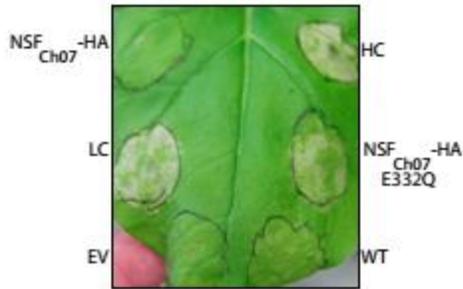


**Fig. S4.**  $\alpha$ -SNAP protein encoded by alternate splicing of the low-copy  $\alpha$ -SNAP transcript does not appreciably accumulate in soybean roots or *N. benthamiana* leaves. (A) Anti- $\alpha$ -SNAP<sub>*Rhg1*LC</sub> immunoblot of three separate samples of agroinfiltrated *N. benthamiana* leaves expressing  $\alpha$ -SNAP WT,  $\alpha$ -SNAP<sub>*Rhg1*LC</sub>,  $\alpha$ -SNAP<sub>*Rhg1*LC<sub>Splice</sub></sub>, or empty vector. Ponceau S staining shows relative protein levels. Immunoblot labels: EV, empty vector; LC,  $\alpha$ -SNAP<sub>*Rhg1*LC</sub>; LC<sub>Splice</sub>,  $\alpha$ -SNAP<sub>*Rhg1*LC<sub>Splice</sub></sub>; WT,  $\alpha$ -SNAP<sub>*Rhg1*WT</sub>. (B) Anti- $\alpha$ -SNAP<sub>*Rhg1*LC</sub> immunoblot of soybean Forrest root lysates, transgenic root lysates from Williams 82 expressing  $\alpha$ -SNAP<sub>*Rhg1*LC</sub>, empty vector, or  $\alpha$ -SNAP<sub>*Rhg1*LC<sub>Splice</sub></sub>, or purified recombinant  $\alpha$ -SNAP<sub>*Rhg1*LC<sub>Splice</sub></sub> to confirm anti- $\alpha$ -SNAP<sub>*Rhg1*LC</sub> recognition of the  $\alpha$ -SNAP<sub>*Rhg1*LC<sub>Splice</sub></sub> protein. Note that a low-abundance band is present in Wm82 transgenic roots agroinfiltrated with  $\alpha$ -SNAP<sub>*Rhg1*LC<sub>Splice</sub></sub> construct but not empty vector.

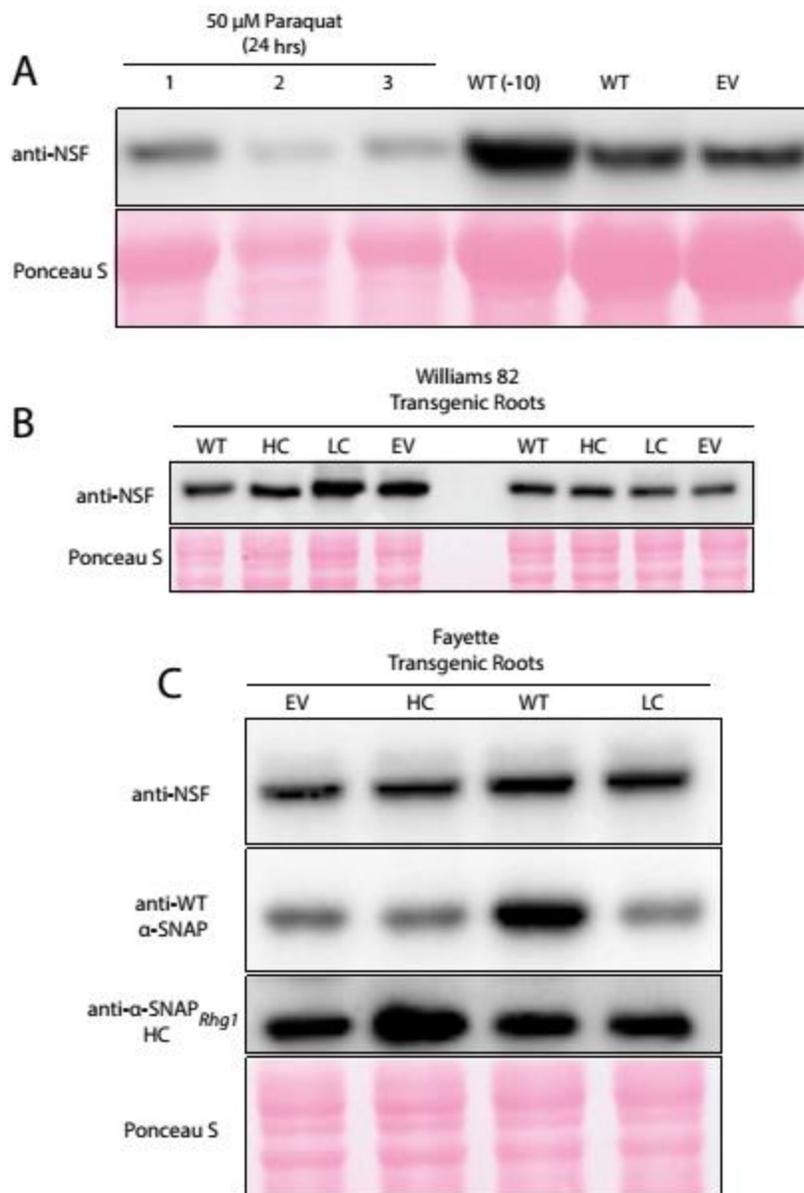


**Fig. S5.** *Rhg1* resistance-type  $\alpha$ -SNAP cytotoxicity is dosage-dependent and occurs independent of the other *Rhg1* locus-encoded genes. (A) *N. benthamiana* leaf agroinfiltrated with native genomic *Rhg1* three-gene blocks (3G Native *Rhg1*) containing *Glyma.18G022400* or *Glyma.18G022700* and the *Glyma.18G022500* alleles encoding the respective single-copy, low-copy, or high-copy *Rhg1*  $\alpha$ -SNAPs. Overexpressed  $\alpha$ -SNAP<sub>*Rhg1*</sub>LC (OX LC) and an empty vector were agroinfiltrated as controls. Cytotoxic symptoms in *N. benthamiana* still occur from expression of  $\alpha$ -SNAPs driven by native soybean *Rhg1* promoters, albeit at a decreased rate and severity compared with expression from a strong ubiquitin promoter. All constructs were

infiltrated at  $OD_{600}$  0.60. An image is shown for 9 d after agroinfiltration. LC,  $\alpha$ -SNAP<sub>RhgI</sub>LC expressed from the soybean ubiquitin promoter. (B) *N. benthamiana* leaf agroinfiltrated with serial two fold dilutions of  $\alpha$ -SNAP<sub>RhgI</sub>LC or an empty vector control. Leaf shown 6 d after agroinfiltration. (C) *N. benthamiana* leaf agroinfiltrated with a 1:3 vs. a 3:1 mixture of  $\alpha$ -SNAP<sub>RhgI</sub>LC and  $\alpha$ -SNAP<sub>RhgI</sub>WT shows further decreases in cytotoxic progression compared with  $\alpha$ -SNAP<sub>RhgI</sub>LC alone. Leaf shown ~8 d after infiltration.

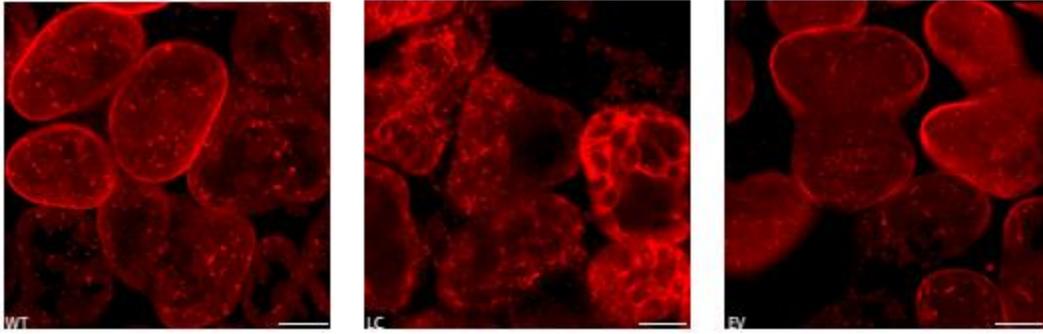


**Fig. S6.** Expression of an NSF lacking ATPase activity phenocopies  $\alpha$ -SNAP *Rhg1* expression and is cytotoxic to *N. benthamiana*. *N. benthamiana* leaf expressing soybean NSF<sub>Ch07</sub>-HA, the ATPase-null NSF<sub>Ch07</sub>-HA (E332Q),  $\alpha$ -SNAP<sub>Rhg1</sub>LC,  $\alpha$ -SNAP<sub>Rhg1</sub>HC,  $\alpha$ -SNAP<sub>Rhg1</sub>WT, or empty vector control at 7 d after agroinfiltration. HC,  $\alpha$ -SNAP<sub>Rhg1</sub>HC. NSF and  $\alpha$ -SNAP expression was from the soybean ubiquitin promoter. NSF-HA<sub>Ch07</sub>E332Q but not WT NSF<sub>Ch07</sub>-HA expression causes cell death similar to  $\alpha$ -SNAP<sub>Rhg1</sub>LC or HC.

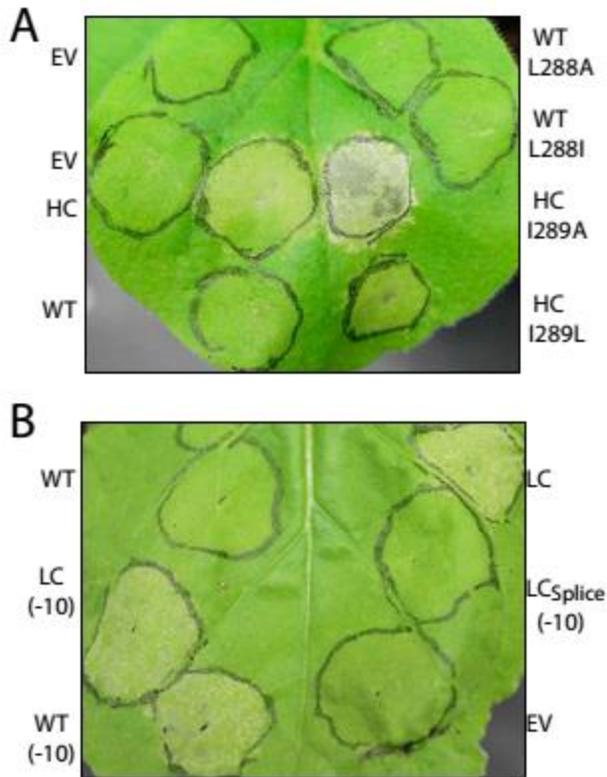


**Fig. S7.** Expression of  $\alpha$ -SNAP<sub>Rhg1</sub>WT(-10) raises NSF levels in *N. benthamiana* leaves, but paraquat treatment of *N. benthamiana* leaves, or transgenic expression of *Rhg1* resistance-type  $\alpha$ -SNAP in soybean hairy roots, does not detectably raise abundance of NSF. (A) Immunoblot of *N. benthamiana* leaf lysates 24 h after infiltrating with 50  $\mu$ M paraquat (methyl viologen) or 3 d after agroinfiltration delivery of the indicated  $\alpha$ -SNAPs. (B) Anti-NSF immunoblots on transgenic Williams 82 root lysates expressing the indicated  $\alpha$ -SNAPs. (C) Anti-NSF

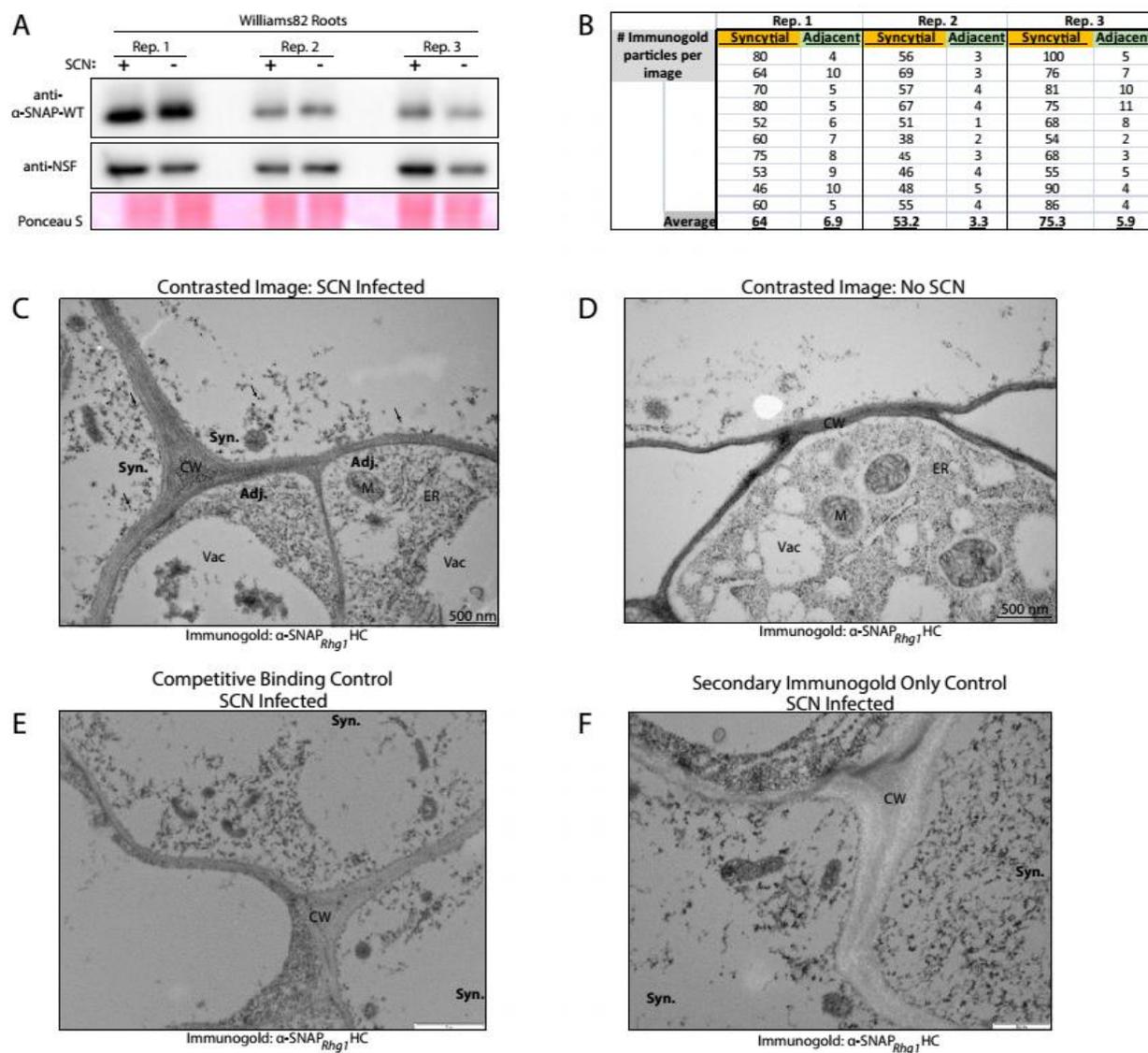
immunoblots on transgenic Fayette root lysates expressing the respective  $\alpha$ -SNAPs. Ponceau S staining shows relative protein levels. WT(-10),  $\alpha$ -SNAP<sub>Rhgl</sub>WT(-10).



**Fig. S8.** Resistance-type  $\alpha$ -SNAP expression appears to disrupt localization of the trans-Golgi network/early endosome marker Syp61-mCherry in *N. benthamiana*. Confocal images of *N. benthamiana* mesophyll cells co-expressing Syp61-mCherry and  $\alpha$ -SNAP<sub>RhgI</sub>WT,  $\alpha$ -SNAP<sub>RhgI</sub>LC, or empty vector. Images are at 3 d after agroinfiltration; n = 20 for each construct. (Scale bars, 20  $\mu$ m.)



**Fig. S9.** Penultimate leucine substitutions of  $\alpha$ -SNAP<sub>RhgI</sub>WT are not macroscopically cytotoxic, but removing the final 10 C-terminal residues is strongly cytotoxic. (A) *N. benthamiana* leaf expressing  $\alpha$ -SNAP<sub>RhgI</sub>HC-I289L or -I289A or  $\alpha$ -SNAP<sub>RhgI</sub>WT-L288I or -L288A shows that substitutions at the penultimate amino acid position influence  $\alpha$ -SNAP<sub>RhgI</sub>HC cytotoxicity but do not confer macroscopic cytotoxicity to  $\alpha$ -SNAP<sub>RhgI</sub>WT. Image shown at ~6 d post agroinfiltration. Respective penultimate residue substitutions are as indicated. (B) *N. benthamiana* leaf expressing  $\alpha$ -SNAP<sub>RhgI</sub>WT truncated at the C terminus causes cell death similar to *RhgI* resistance-type  $\alpha$ -SNAPs. Agroinfiltrated constructs were LC,  $\alpha$ -SNAP<sub>RhgI</sub>LC; LC(-10),  $\alpha$ -SNAP<sub>RhgI</sub>LC(-10); LC<sub>splice</sub>(-10),  $\alpha$ -SNAP<sub>RhgI</sub>LC<sub>splice</sub>(-10); WT,  $\alpha$ -SNAP<sub>RhgI</sub>WT; WT(-10),  $\alpha$ -SNAP<sub>RhgI</sub>WT(-10); and empty vector.



**Fig. S10.** Quantification of *Rhg1*  $\alpha$ -SNAPs in developing syncytia and confirmation of  $\alpha$ -SNAP<sub>Rhg1</sub>HC specificity when used in immunogold labeling of electron microscopy sections of SCN-infested roots. (A) Immunoblot of Williams 82 tissue samples from SCN-infested root regions harvested 4 d after SCN infection. Blots were probed with the indicated antibodies. (B) Number of  $\alpha$ -SNAP<sub>Rhg1</sub>HC immunogold particles detected in syncytial cells vs. adjacent cells in SCN-infested Fayette roots. Data from three independent experiments are shown. (C) Contrasted electron micrograph of the syncytium and adjacent cell of Fayette root infested with SCNs, after

immunogold label detection using anti- $\alpha$ -SNAP<sub>RhgI</sub>HC primary antibody [similar to Fig. 4C (non-contrasted)]. Adj., adjacent cell; CW, cell wall; ER, endoplasmic reticulum; M, mitochondrion; Syn., syncytial cell; Vac, vacuole. Arrows highlight four of many gold particle-labeled  $\alpha$ -SNAP<sub>RhgI</sub>HC regions. (D) Contrasted electron micrograph of mock-inoculated Fayette root after immunogold label detection using anti- $\alpha$ -SNAP<sub>RhgI</sub>HC primary antibody. (E) Electron micrograph of a syncytium site of Fayette root infested with SCNs, where the primary anti- $\alpha$ -SNAP<sub>RhgI</sub>HC antibody was competitively bound with a 10-fold molar excess of antigen (recombinant  $\alpha$ -SNAP<sub>RhgI</sub>HC protein) before immunolabeling of the microscopy section. After the initial competitive binding, anti- $\alpha$ -SNAP<sub>RhgI</sub>HC primary antibody was incubated with fixed cross-sections of SCN-infested Fayette roots and probed with secondary goat anti-rabbit antibody conjugated to 15-nm gold particles. Multiple cross-sections were examined using competitively bound  $\alpha$ -SNAP<sub>RhgI</sub>HC primary antibody and little to no gold particle labeling was observed, indicating high antigen specificity. (F) Immunogold labeling using only secondary goat anti-rabbit antibody on SCN-infested roots. No previous incubations with  $\alpha$ -SNAP<sub>RhgI</sub>HC antibody were performed. Little to no gold particle labeling is present, indicating  $\alpha$ -SNAP<sub>RhgI</sub>HC labeling in SCN-infested roots is highly specific. (Scale bars in E and F, 1  $\mu$ m.)

## 4.10 Supporting Tables

Table S1. Oligonucleotides used

Name	Sequence, 5'-3'
SuNSF 7 Rev	GTGGCGGCGCTCTATTATAACCTAACACATCCTGGAGGCAATCATA
ExpV For	TAA TAG AGC GGC CGC CAC C
SuNSF 13 Rev	GTGGCGGCGCTCTATTATCTAACACATCCTGGAGGCAATCATAG
SuNSF 7 For	CGC GAA CAG ATT GGA GGT GCG AGT CGG TTC GGG TTA TC
SuNSF 13 For	CGCGAACAGATTGGAGGTTTCGGCTTATCGTCTTCGTCTTCCTC
ExpV Rev	ACC TCC AAT CTG TTC GCG GTG
NSF 07g SUMO Exp For	CGC GAA CAG ATT GGA GGTGCGAGTCGGTTCGGGTATC
NSF 07g SUMO Exp Rev	gtggcgccgctctattaTAACCTAACACATCCTGGAGGCAATCATG
NSF 13 cDNA spec Rev	GGTCATTACAGTTTGAGAGCAGCAC
NSF 13 cDNA spec For	GCCAAAGAACAGAGAACATAGAGGC
NSF 07 E332Q For	cAAATTGATGCTATTGCAAGTCAAGAGGTTTC
NSF 07 E332Q Rev	CATCTCGAGTTGAACCTCTTGACTTG
NSF 07g cDNA For	ATG GCG AGT CGG TTC GGG TTA TCG T
NSF 07g cDNA Rev	TAA CCT AAC AAC ATC CTG GAG GCA ATC ATA GAA ATG AGC
NSF 07g SUMO Exp For	CGC GAA CAG ATT GGA GGTGCGAGTCGGTTCGGGTATC
NSF 07g SUMO Exp Rev	gtggcgccgctctattaTAACCTAACACATCCTGGAGGCAATCAT
pRham 2590 Fuse Will Rev	GTGGCGGCGCTCTATTAAAGTAAGATCATCCTCCTCAAGTTCCTTG
pRham 2590 Fuse For	CGCGAACAGATTGGAGGTGCCGATCAGTTATCGAAGGGA GAG G
pRham 2590 Fuse Pek Rev	gtggcgccgctctattaagtaataacctcactcctcaagttctttgg
WT aSNAP C-10 Trunc Rev	tcaTTTCAGCTTTTCCTTCACCCTTAAGAGa
WT aSNAP C-10 Trunc For	GAAAAGCTGAAATGATGAATTGTACCTTTAATATTCTGGTGGTTGG
Ch 09 aSNAP cDNA For	GTGTTGGCAAAGGGTGATGAC
Ch 09 aSNAP cDNA Rev	CAAAGCTGAGAGTAACCTAATTGGCAG
Ch 02 aSNAP CDNA For	TTCCAATATGGCGATCATTGG
Ch 02 aSNAP CDNA Rev	ACCGAAAGAAGACCATGGTGC
Ch 11 aSNAP CDNA For	CGATCAATCCATCCATCTTCACTTGC
Ch 11 aSNAP CDNA Rev	CAAACAATAGGTCCAACCGCCAG
Ch 11 aSNAP PIPE Rev	AATTGCGCCTTTTCAAGTAAGATCATCCTCCTCAAGTTCCTTTGG
Ch 11 aSNAP PIPE For	TTGTTGACTCGACAGATGGCCGATCAGTTATCCAAAGG
Ch 09 aSNAP PIPE For	ttgTTGACTCGACAGATGCTTGTGGCCCTTGTTCG
Ch 09 aSNAP PIPE Rev	AATTGCGCCTTTTCAAGTAAGATCATCCTCCTCAAGTTCCTTTAC
Ch 02 aSNAP PIPE For	TTGTTGACTCGACAGATGGCCGATCATTGGCCAG
Ch 02 aSNAP PIPE Rev	AATTGCGCCTTTTCAAGTAAGATCATCCTCCTCGATTTCTTTG
pBS ter PIPE For	TGAAAAGGGCGAATTCGACCC
pBS Gmubi PIPE Rev	CTGTGAGTCAACAATCACAGATAAATC
For aSNAP HC PIPE, 289 Ala	GCT GCT ACT TGA TAA TAG AGC GGC CGC CA
Rev aSNAP HC PIPE, 289 Leu	AGT AAG AGC CTC ATG CTG CTC AAG TTC TTT GGC
For aSNAP HC PIPE, 289 Leu	TGAGCAGCATGAGGCTTACTTGAAACCCAGCTTCTT GTA CAA AG
Rev aSNAP HC PIPE, 289 ala	AGT AGC AGC CTC ATG CTG CTC AAG TTC TTT GGC
For aSNAP WT PIPE, L288A	TGAGGAGGATGATGCTACTTGAAACCCAGCTTCTTGTGA CAA AG
Rev aSNAP WT PIPE, L288A	AGT AGC ATC ATC CTC CTC AAG TTC TTT GGC
Rev aSNAP WT PIPE, L288I	AGT AAT ATC ATC CTC CTC AAG TTC TTT GGC
aSNAP Rhg1 LC	ctctgtaaagaggaggtgtgtgctat
aSNAP Rhg1 LC Rev qPCR	gcaatgtccgccaacaate
aSNAP Rhg1 LC Splice	gtaaagaggaggaaactggatcc
aSNAP LC cDNA Rev	AGTAATAACCTCATACTCCTCAAGTT
WT aSNAP cDNA Rev	AGTAAGATCATCCTCCTCAAGTTC
SUMO aSNAP WT Rev	tcaTTTCAGCTTTTCCTTCACCCTTAAGAGa
SUMO aSNAP WT C-10 Trunc For	gaaaagctgaaatgatgaattgtacctttaatattcctgggtggtgg
pRham aSNAP WT Rev	gtg gcg gcc gct cta tta agt aag atc atc ctc ctc aag ttc ttt gg
pRham aSNAP For	cgc gaa cag att gga ggt gcc gat cag tta tcg aag gga gag g
pRham 2590 Fuse Peking Rev	gtg gcg gcc gct cta tta agt aat aac ctc ata ctc ctc aag ttc ttt gg
pRham 2590 Fuse Fay Rev	gtg gcg gcc gct cta tta agt aat agc ctc atg ctg ctc aag ttc

For, forward; qPCR, quantitative PCR; Rev, reverse.

Table S1

#### 4.11 References

- Asrat S, de Jesus DA, Hempstead AD, Ramabhadran V, Isberg RR** (2014) Bacterial pathogen manipulation of host membrane trafficking. *Annu Rev Cell Dev Biol* **30**: 79-109
- Barnard RJ, Morgan A, Burgoyne RD** (1996) Domains of alpha-SNAP required for the stimulation of exocytosis and for N-ethylmaleimide-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
- Barszczewski M, Chua JJ, Stein A, Winter U, Heintzmann R, Zilly FE, Fasshauer D, Lang T, Jahn R** (2008) A novel site of action for alpha-SNAP in the SNARE conformational cycle controlling membrane fusion. *Mol Biol Cell* **19**: 776-784
- Bassham DC, Raikhel NV** (1999) The pre-vacuolar t-SNARE AtPEP12p forms a 20S complex that dissociates in the presence of ATP. *Plant J* **19**: 599-603
- Batoko H, Zheng HQ, Hawes C, Moore I** (2000) A rab1 GTPase is required for transport between the endoplasmic reticulum and golgi apparatus and for normal golgi movement in plants. *Plant Cell* **12**: 2201-2218
- Bekal S, Domier LL, Gonfa B, Lakhssassi N, Meksem K, Lambert KN** (2015) A SNARE-Like Protein and Biotin Are Implicated in Soybean Cyst Nematode Virulence. *PLoS One* **10**: e0145601
- Chae TH, Kim S, Marz KE, Hanson PI, Walsh CA** (2004) The *hyh* mutation uncovers roles for alpha Snap in apical protein localization and control of neural cell fate. *Nat Genet* **36**: 264-270
- Colgrove AL, Niblack TL** (2008) Correlation of Female Indices From Virulence Assays on Inbred Lines and Field Populations of *Heterodera glycines*. *J Nematol* **40**: 39-45
- Collins NC, Thordal-Christensen H, Lipka V, Bau S, Kombrink E, Qiu JL, Huckelhoven R, Stein M, Freialdenhoven A, Somerville SC, Schulze-Lefert P** (2003) SNARE-protein-mediated disease resistance at the plant cell wall. *Nature* **425**: 973-977
- Concibido VC, Diers BW, Arelli PR** (2004) A Decade of QTL Mapping for Cyst Nematode Resistance in Soybean. *Crop Sci.* **44**: 1121-1131
- Cook DE, Bayless AM, Wang K, Guo X, Song Q, Jiang J, Bent AF** (2014) Distinct Copy Number, Coding Sequence, and Locus Methylation Patterns Underlie Rhg1-Mediated Soybean Resistance to Soybean Cyst Nematode. *Plant Physiol* **165**: 630-647

- 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
- Crooks GE, Hon G, Chandonia JM, Brenner SE** (2004) WebLogo: a sequence logo generator. *Genome Res* **14**: 1188-1190
- Dalal S, Rosser MF, Cyr DM, Hanson PI** (2004) Distinct roles for the AAA ATPases NSF and p97 in the secretory pathway. *Mol Biol Cell* **15**: 637-648
- Dangl JL, Horvath DM, Staskawicz BJ** (2013) Pivoting the plant immune system from dissection to deployment. *Science* **341**: 746-751
- Davis EL, Hussey RS, Mitchum MG, Baum TJ** (2008) Parasitism proteins in nematode-plant interactions. *Curr Opin Plant Biol* **11**: 360-366
- de Almeida Engler J, Gheysen G** (2013) Nematode-induced endoreduplication in plant host cells: why and how? *Mol Plant Microbe Interact* **26**: 17-24
- Donald PA, Pierson PE, St Martin SK, Sellers PR, Noel GR, Macguidwin AE, Faghihi J, Ferris VR, Grau CR, Jardine DJ, Melakeberhan H, Niblack TL, Stienstra WC, Tylka GL, Wheeler TA, Wysong DS** (2006) Assessing *Heterodera glycines*-Resistant and Susceptible Cultivar Yield Response. *J Nematol* **38**: 76-82
- Elguero E, Delicat-Loembet LM, Rougeron V, Arnathau C, Roche B, Becquart P, Gonzalez JP, Nkoghe D, Sica L, Leroy EM, Durand P, Ayala FJ, Ollomo B, Renaud F, Prugnolle F** (2015) Malaria continues to select for sickle cell trait in Central Africa. *Proc Natl Acad Sci U S A* **112**: 7051-7054
- Gheysen G, Mitchum MG** (2011) How nematodes manipulate plant development pathways for infection. *Curr Opin Plant Biol* **14**: 415-421
- Gu Y, Innes RW** (2011) The KEEP ON GOING protein of *Arabidopsis* recruits the ENHANCED DISEASE RESISTANCE1 protein to trans-Golgi network/early endosome vesicles. *Plant Physiol* **155**: 1827-1838
- Hanson PI, Roth R, Morisaki H, Jahn R, Heuser JE** (1997) Structure and conformational changes in NSF and its membrane receptor complexes visualized by quick-freeze/deep-etch electron microscopy. *Cell* **90**: 523-535
- Hoefle C, Huckelhoven R** (2008) Enemy at the gates: traffic at the plant cell pathogen interface. *Cell Microbiol* **10**: 2400-2407
- Inada N, Ueda T** (2014) Membrane trafficking pathways and their roles in plant-microbe interactions. *Plant Cell Physiol* **55**: 672-686

- Jahn R, Scheller RH** (2006) SNAREs--engines for membrane fusion. *Nat Rev Mol Cell Biol* **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 Physiol* **155**: 1960-1975
- Kang BC, Yeam I, Jahn MM** (2005) Genetics of plant virus resistance. *Annu Rev Phytopathol* **43**: 581-621
- Klock HE, Lesley SA** (2009) The Polymerase Incomplete Primer Extension (PIPE) method applied to high-throughput cloning and site-directed mutagenesis. *Methods Mol Biol* **498**: 91-103
- Kyndt T, Vieira P, Gheysen G, de Almeida-Engler J** (2013) Nematode feeding sites: unique organs in plant roots. *Planta* **238**: 807-818
- Lambert KN, Bekal S, Domier LL, Niblack TL, Noel GR, Smyth CA** (2005) Selection of *Heterodera glycines* chorismate mutase-1 alleles on nematode-resistant soybean. *Mol Plant Microbe Interact* **18**: 593-601
- Lee TG, Kumar I, Diers BW, Hudson ME** (2015) Evolution and selection of Rhg1, a copy-number variant nematode-resistance locus. *Mol Ecol* **24**: 1774-1791
- 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
- Naydenov NG, Harris G, Brown B, Schaefer KL, Das SK, Fisher PB, Ivanov AI** (2012) Loss of soluble N-ethylmaleimide-sensitive factor attachment protein alpha (alphaSNAP) induces epithelial cell apoptosis via down-regulation of Bcl-2 expression and disruption of the Golgi. *J Biol Chem* **287**: 5928-5941
- Niblack T, Colgrove A, Colgrove K, Bond J** (2008) Shift in virulence of soybean cyst nematode is associated with use of resistance from PI 88788. *Plant Health Prog* **10**
- Niblack TL, Lambert KN, Tylka GL** (2006) A model plant pathogen from the kingdom Animalia: *Heterodera glycines*, the soybean cyst nematode. *Annu Rev Phytopathol* **44**: 283-303
- Niks RE, Qi X, Marcel TC** (2015) Quantitative resistance to biotrophic filamentous plant pathogens: concepts, misconceptions, and mechanisms. *Annu Rev Phytopathol* **53**: 445-470

- Pant SR, Krishnavajhala A, McNeece BT, Lawrence GW, Klink VP** (2015) The syntaxin 31-induced gene, LESION SIMULATING DISEASE1 (LSD1), functions in Glycine max defense to the root parasite *Heterodera glycines*. *Plant Signal Behav* **10**: e977737
- Pier GB, Grout M, Zaidi T, Meluleni G, Mueschenborn SS, Banting G, Ratcliff R, Evans MJ, Colledge WH** (1998) *Salmonella typhi* uses CFTR to enter intestinal epithelial cells. *Nature* **393**: 79-82
- Rancour DM, Dickey CE, Park S, Bednarek SY** (2002) Characterization of AtCDC48. Evidence for multiple membrane fusion mechanisms at the plane of cell division in plants. *Plant Physiol* **130**: 1241-1253
- Ryu JK, Min D, Rah SH, Kim SJ, Park Y, Kim H, Hyeon C, Kim HM, Jahn R, Yoon TY** (2015) Spring-loaded unraveling of a single SNARE complex by NSF in one round of ATP turnover. *Science* **347**: 1485-1489
- Sanyal S, Krishnan KS** (2001) Lethal comatose mutation in *Drosophila* reveals possible role for NSF in neurogenesis. *Neuroreport* **12**: 1363-1366
- Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A** (2012) Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**: 676-682
- Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, 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. *Nature* **463**: 178-183
- Song J, Keppler BD, Wise RR, Bent AF** (2015) PARP2 Is the Predominant Poly(ADP-Ribose) Polymerase in *Arabidopsis* DNA Damage and Immune Responses. *PLoS Genet* **11**: e1005200
- Song QX, Liu YF, Hu XY, Zhang WK, Ma B, Chen SY, Zhang JS** (2011) Identification of miRNAs and their target genes in developing soybean seeds by deep sequencing. *BMC Plant Biol* **11**: 5
- Uemura T, Kim H, Saito C, Ebine K, Ueda T, Schulze-Lefert P, Nakano A** (2012) Qa-SNAREs localized to the trans-Golgi network regulate multiple transport pathways and extracellular disease resistance in plants. *Proc Natl Acad Sci U S A* **109**: 1784-1789
- Vivona S, Cipriano DJ, O'Leary S, Li YH, Fenn TD, Brunger AT** (2013) Disassembly of all SNARE complexes by N-ethylmaleimide-sensitive factor (NSF) is initiated by a

conserved 1:1 interaction between alpha-soluble NSF attachment protein (SNAP) and SNARE complex. *J Biol Chem* **288**: 24984-24991

**Wickner W, Schekman R** (2008) Membrane fusion. *Nat Struct Mol Biol* **15**: 658-664

**Yu A, Lepere G, Jay F, Wang J, Bapaume L, Wang Y, Abraham AL, Penterman J, Fischer RL, Voinnet O, Navarro L** (2013) Dynamics and biological relevance of DNA demethylation in Arabidopsis antibacterial defense. *Proc Natl Acad Sci U S A* **110**: 2389-2394

**Zhao C, Slevin JT, Whiteheart SW** (2007) Cellular functions of NSF: not just SNAPs and SNAREs. *FEBS Lett* **581**: 2140-2149

**Zhao M, Wu S, Zhou Q, Vivona S, Cipriano DJ, Cheng Y, Brunger AT** (2015) Mechanistic insights into the recycling machine of the SNARE complex. *Nature* **518**: 61-67

**Zick M, Orr A, Schwartz ML, Merz AJ, Wickner WT** (2015) Sec17 can trigger fusion of trans-SNARE paired membranes without Sec18. *Proc Natl Acad Sci U S A* **112**: E2290-2297

## **Chapter 5: An atypical NSF (N-ethylmaleimide Sensitive Factor) enables the viability of nematode-resistant *Rhg1* soybeans**

The material in this chapter is currently being prepared for publication.

Authors: Adam M. Bayless, Ryan Zapotocny, Derrick Grunwald, Brian Diers, Andrew F. Bent

Contributions: I conceived of the project with Andrew Bent, and performed the majority of the experiments. Ryan Zapotocny performed bioinformatic analyses, Derrick Grunwald assisted with RAN07 detection screens in soybeans and Brian Diers contributed soybean cross and mapping population data. Adam Bayless and Andrew Bent wrote the manuscript.

## 5.1 Abstract

In eukaryotes, sustained vesicle trafficking requires physical interaction of NSF and  $\alpha$ -SNAP proteins (N-ethylmaleimide Sensitive Factor and  $\alpha$ -Soluble NSF Attachment Protein) to disassemble SNARE (soluble NSF attachment protein receptor) vesicle docking protein complexes. The Resistance to *Heterodera glycines* 1 (*Rhg1*) locus of soybean (*Glycine max*) confers resistance to soybean cyst nematode (SCN), the most yield-damaging pathogen of soybean. *Rhg1* loci encode multiple repeat copies of atypical  $\alpha$ -SNAP proteins, which unlike wild-type (WT)  $\alpha$ -SNAPs, are defective in binding NSF and cytotoxic in certain contexts. Here, we show that presence of certain *Rhg1* soybean haplotypes is associated with selective loss of functional unlinked WT  $\alpha$ -SNAP genes. Moreover, we discovered an unusual NSF allele in all *Rhg1*-containing soybean lines that encodes five N-domain amino acid polymorphisms. We termed this *Glyma.07g195900* allele  $NSF_{\text{RAN07}}$  (*Rhg1*-associated NSF on chromosome 07). Modeling of  $NSF_{\text{RAN07}}$  to structures of mammalian NSF/ $\alpha$ -SNAP/SNARE complexes indicates that at least three of these  $NSF_{\text{RAN07}}$  polymorphisms map to the  $\alpha$ -SNAP binding interface. *In vitro* binding assays demonstrate that  $NSF_{\text{RAN07}}$  has stronger binding to resistance-associated *Rhg1*  $\alpha$ -SNAPs than does wild-type  $NSF_{\text{Ch07}}$  protein. Co-expression of the  $NSF_{\text{RAN07}}$  protein is more protective against *Rhg1*  $\alpha$ -SNAP cytotoxicity than wild-type  $NSF_{\text{Ch07}}$  protein. Investigation of the previously reported severe segregation distortion between *Rhg1* and an unlinked locus on chromosome 07, which includes the *Glyma.07g19590* NSF gene, revealed strict co-presence of disease resistance *Rhg1* alleles and the  $NSF_{\text{RAN07}}$  allele across 855 soybean accessions, and in all examined *Rhg1*<sup>+</sup> progeny from biparental crosses, confirming the functional necessity of this co-occurrence. Hence co-evolution of multiple components of the core vesicular trafficking machinery in *Rhg1*-containing soybeans has balanced the acquisition of

an otherwise toxic housekeeping protein that confers an important disease resistance trait. Our findings also indicate that efforts to engineer *Rhg1*-related mechanisms for cyst nematode resistance in soybean or other plant species will require a compatible NSF protein partner for the resistance-conferring  $\alpha$ -SNAP.

## 5.2. Introduction

Soybean cyst nematode (*Heterodera glycines*; SCN) is consistently the most damaging disease of U.S. soybeans, one of the world's main food crops (Niblack et al., 2006; Jones et al., 2013; Allen et al., 2016; Mitchum, 2016; T. W. Allen, 2017). Plant parasitic nematodes, including cyst nematodes, infest the roots of many valuable crops and establish elaborate feeding structures (Kyndt et al., 2013). Cyst nematodes secrete a complex arsenal of effector molecules that modulate the host's physiology and promote the fusion of neighboring host cells into a large unicellular feeding site, termed a syncytium (Gheysen and Mitchum, 2011; Hewezi and Baum, 2013; Mitchum et al., 2013). The soybean *Rhg1* (Resistance to *Heterodera glycines*) locus has been very widely used by soybean breeders and growers as the best available disease resistance locus to reduce the damage caused by SCN (Concibido et al., 2004; Mitchum, 2016). The relevant genes at *Rhg1* do not encode proteins normally associated with disease resistance (Mitchum et al., 2004; Dodds and Rathjen, 2010 ; Cook et al., 2012; Lee et al., 2015). Instead, resistance is mediated by copy number variation of multiple genes at the *Rhg1* locus, one of which encodes a protein with high similarity to known  $\alpha$ -SNAP proteins (Cook et al., 2012; Cook et al., 2014; Lee et al., 2015).  $\alpha$ -SNAP (alpha-Soluble NSF-Attachment Protein) is a ubiquitous housekeeping protein in plants and animals that facilitates cellular vesicular trafficking by mediating the disassembly and reuse of the four-protein bundles of SNARE proteins (soluble NSF attachment protein receptor proteins) that form when t-SNARE and v-SNARE proteins anneal during vesicle docking to target membranes (Jahn and Scheller, 2006; Baker and Hughson, 2016; Zhao and Brunger, 2016).  $\alpha$ -SNAP functions together with the ATPase NSF (N-ethylmaleimide Sensitive Factor) to carry out this SNARE bundle disassembly (Zhao and Brunger, 2015). We recently discovered that the soybean resistance-associated  $\alpha$ -

SNAPs encoded by *Rhg1* are defective proteins that bind less well to NSF and to SNARE/NSF complexes, and which disrupt vesicle trafficking *in planta* (Bayless et al., 2016). SCN-resistant soybeans carry genes at other loci that encode wild-type  $\alpha$ -SNAPs, which can functionally complement the *Rhg1* encoded resistance-type  $\alpha$ -SNAPs. However, the relative abundance of *Rhg1*-encoded defective  $\alpha$ -SNAP variants increases substantially within host syncytium cells at the nematode feeding site (Bayless et al., 2016). Apparently, soybeans use these aberrant  $\alpha$ -SNAPs to alter the compatibility of soybean cells at the nematode feeding site, and thereby disrupt nematode growth and reproduction.

The complex *Rhg1* locus on soybean chromosome 18 is a tandemly repeated block of four genes: *Glyma.18G022400* (formerly *Glyma18g02580*), *Glyma.18G022500* (formerly *Glyma18g02590*), *Glyma.18G022600* (formerly *Glyma18g02600*) and *Glyma.18G022700* (formerly *Glyma18g02610*). SCN-susceptible soybeans carry only a single copy of the above four genes, and encode an  $\alpha$ -SNAP matching the wild-type (WT)  $\alpha$ -SNAP consensus which maintains normal NSF interactions (Cook et al., 2012; Cook et al., 2014; Bayless et al., 2016). However, the resistance-conferring *Rhg1* loci group into two structural classes based on the type of  $\alpha$ -SNAP polymorphisms that they encode, which also correlates perfectly with the copy-number of *Rhg1* repeats that are present, as observed across numerous soybean accessions (Cook et al., 2014; Lee et al., 2015). *Rhg1<sub>HC</sub>* (high copy) loci carry four or more and frequently nine or ten *Rhg1* repeats, and *Rhg1<sub>LC</sub>* (low-copy) loci carry three or fewer *Rhg1* repeats. *Rhg1<sub>LC</sub>* is also known as *rhg1-a* and *Rhg1<sub>HC</sub>* is also known as *rhg1-b*. *Rhg1<sub>HC</sub>* and *Rhg1<sub>LC</sub>* encode similar yet distinct  $\alpha$ -SNAP variants that are impaired in normal  $\alpha$ -SNAP-NSF interactions (Bayless et al., 2016). All *Rhg1<sub>HC</sub>* loci examined to date also have one *Rhg1* repeat that encodes a WT  $\alpha$ -SNAP along with multiple repeats encoding a resistance-type  $\alpha$ -SNAP, while *Rhg1<sub>LC</sub>* loci encode only

resistance-type  $\alpha$ -SNAPs and no WT  $\alpha$ -SNAP (Cook et al., 2012; Cook et al., 2014; Lee et al., 2015). Plants carrying *Rhg1<sub>HC</sub>* or *Rhg1<sub>LC</sub>* loci exhibit elevated transcript abundance that correlates approximately with copy number for the repeat genes, including the *Rhg1*  $\alpha$ -SNAP gene (Cook et al., 2012; Cook et al., 2014). In experiments performed in *N. benthamiana* leaves, high expression of these resistance-conferring  $\alpha$ -SNAPs hindered vesicular trafficking and eventually elicited cell death, but co-expression of wild type soybean  $\alpha$ -SNAPs diminished this cytotoxicity (Bayless et al., 2016). These findings, together with the relative protein abundance findings noted in the previous paragraph, suggest that at least one core mechanism of *Rhg1*-mediated SCN resistance is modulation of vesicle trafficking and cell health at the SCN feeding site.

Two other genes within the *Rhg1* repeat were indicated (or reported) by Cook et al. to contribute to *Rhg1<sub>HC</sub>*-mediated SCN resistance (Cook et al., 2012). *Glyma.18G022400* encodes an apparent amino acid permease and *Glyma.18G022700* encodes a wound-inducible protein otherwise lacking annotated domains or predicted functions, but their molecular function in SCN resistance remains unknown. Liu et al. recently provided evidence that the *Rhg1<sub>LC</sub>*  $\alpha$ -SNAP may function differently than the *Rhg1<sub>HC</sub>*  $\alpha$ -SNAP (Liu et al., 2017).

The eukaryotic endomembrane network is an intricate sorting and secretion system that ferries cargoes between cellular compartments using transport vesicles. Cognate SNARE proteins on the surface of vesicle and target membranes drive membrane fusion by “zippering” into stable bundles (SNARE complexes), which pull the membranes together (Jahn and Scheller, 2006; Wickner and Schekman, 2008). SM (Sec1/Munc18) family proteins are essential for vesicle fusion *in vivo*, and reportedly enhance fusion by guiding the pairing of particular SNARE subsets at specific endomembrane compartments (Sudhof and Rothman, 2009; Lobingier et al.,

2014; Baker and Hughson, 2016). The role of  $\alpha$ -SNAP and NSF as dedicated SNARE chaperones that form the SNARE recycling machinery has been studied extensively (Jahn and Scheller, 2006; Wickner and Schekman, 2008; Wickner, 2010; Zhao et al., 2015; Zick et al., 2015). NSF is a AAA family protein (ATPases Associated with various cellular Activities) containing three well defined domains: the N-domain, which mediates interactions with the  $\alpha$ -SNAP co-chaperone, the D1 ATPase domains, which couples ATP hydrolysis to force-generating conformational changes that remodel SNARE complexes, and the D2 ATPase domain, which mediates NSF hexamerization (Whiteheart et al., 2001; Hanson and Whiteheart, 2005; Zhao et al., 2010).  $\alpha$ -SNAP proteins are required by NSF to co-chaperone SNARE remodeling.  $\alpha$ -SNAP serves both as an adaptor for NSF binding to SNARE complexes and as a stimulator of NSF D1 domain ATPase activity that powers SNARE remodeling/recycling. Beyond this paradigmatic role in disassembling SNARE complexes to sustain pools of free SNARE proteins as acceptors for vesicle fusion, additional roles of  $\alpha$ -SNAP and NSF have been reported, including recent evidence of binding to trans-SNARE complexes to accelerate fusion (Song et al., 2017), as well as binding of channels and other receptors and regulation of apoptosis (Whiteheart and Matveeva, 2004; Zhao et al., 2007; Naydenov et al., 2012; Miao et al., 2013; Zick et al., 2015). The structure and function of  $\alpha$ -SNAP, NSF and SNARE proteins has been elucidated in substantial detail, including cryo-EM structures for 20S complexes that consist of a four-protein SNARE bundle, four  $\alpha$ -SNAPs and six NSFs in various stages of active binding and disassembly (Zhao et al., 2015; Zhao and Brunger, 2016).

Although many animal genomes carry a single NSF and a single  $\alpha$ -SNAP gene, polyploidization and other events have caused most plant genomes to encode multiple NSF and  $\alpha$ -SNAP genes. There are two unlinked NSF genes in soybean. The reference Williams 82

soybean genome (Schmutz et al., 2010) encodes seven Soluble NSF Attachment Protein (SNAP) family members; five putative  $\alpha$ -SNAPs and 2 putative  $\gamma$ -SNAPs. As in animals, plants contain >100 genes encoding diverse SNARE and SNARE-like proteins (Sanderfoot et al., 2000; Jahn and Scheller, 2006). Unlike plant SNARE proteins, NSF and SNAP proteins (not to be confused with the similarly named SNAREs, synaptosomal-associated protein 25 (SNAP25) or soluble N-ethylmaleimide-sensitive factor adaptor protein 33 (SNAP33)) remain poorly studied in plants, with only a handful of reports that go beyond gene identification, and mostly focusing on *Rhg1* and soybean disease resistance (Bassham and Raikhel, 1999; Bachem et al., 2000; Rancour et al., 2002; Matsye et al., 2012; Liu et al., 2017) .

Close analysis of recombinant-inbred lines has shown that a gene at or linked to the soybean Chromosome 11 locus encoding an  $\alpha$ -SNAP makes a minor contribution to SCN resistance in the Peking (*Rhg1<sub>LC</sub>* + *Rhg4*) genetic background (Lakhssassi et al., 2017). Other previous work (Matsye et al., 2012) had identified an allele encoding a splice-variant  $\alpha$ -SNAP in this genetic background, and showed that overexpression of that allele elevated the SCN resistance of transgenic soybean roots, although that work identified it as an allele of the Chromosome 18 *Rhg1* locus despite it now being known to be a Chromosome 11  $\alpha$ -SNAP allele (Cook et al., 2014; Lakhssassi et al., 2017) (and present work). In the present study, we demonstrate that evolution/selection of both *Rhg1<sub>LC</sub>* and this Chromosome 11  $\alpha$ -SNAP gene *Glyma.11G234500* has had major impacts on the relative abundance of WT  $\alpha$ -SNAP proteins in soybeans expressing this type of SCN resistance.

In the present study we also examined soybean NSF proteins, because of our previous discoveries about the atypical *Rhg1*  $\alpha$ -SNAPs (Cook et al., 2014; Bayless et al., 2016). We found that an unusual NSF protein, unlike that encoded in the soybean Williams 82 reference

genome, or any publicly available plant NSF sequence, is encoded by the *Glyma.07g195900* allele present in *Rhg1*-containing lines. We went on to discover that this variant *NSF<sub>RAN07</sub>* (*Rhg1*-associated NSF on Chromosome 07) protein contains unique N-domain polymorphisms which improve upon wild-type NSF in mitigating the cytotoxicity and poor NSF binding activity of SCN resistance-conferring *Rhg1*  $\alpha$ -SNAPs. We then noted that the genetic locus encoding this NSF and neighboring genes has been identified in previous SCN resistance mapping studies, including studies showing strong co-segregation of *Rhg1* and this locus (Webb et al., 1995; Kopisch-Obuch and Diers, 2006; Vuong et al., 2015). Genomewide studies to map loci controlling SCN resistance have most consistently identified *Rhg1*, however, an SCN resistance quantitative trait locus (QTL) on soybean Chromosome 7 (formerly known as Linkage Group (LG) M) was imprecisely mapped as early as 1995 (Webb et al., 1995; Kopisch-Obuch and Diers, 2006). A QTL at a similar location was identified, typically with a lower phenotypic effect than *Rhg1*, in some but not all other SCN resistance QTL mapping studies. This recently included identification of a high-resolution candidate gene interval by Vuong and colleagues (Vuong et al., 2015). They conducted a GWAS study that used the SoySNP50K iSelect BeadChip resource on a core set of 553 soybean accessions and identified an 80 kb candidate gene interval, with the most significant SNP marker residing within the *Glyma.07G195900* gene encoding NSF (Vuong et al., 2015). Returning to the 1995 study of Webb and colleagues, a striking observation regarding this LG-M QTL was its co-segregation with *Rhg1*, an unlinked locus on soybean LG-G (now known as Chromosome 18)(Webb et al., 1995). 91 of 96 lines that had a resistant parent marker type linked to *Rhg1* also had a resistant parent marker type at the LG-M QTL (Webb et al., 1995). This result and subsequent findings of segregation distortion at the Chromosome 7 locus had remained unexplained. We therefore went beyond studies of the

$NSF_{RAN07}$  protein by carrying out an extensive screen of soybean germplasm genotype data, and close examination of F2-derived soybean lines from the soybean NAM (nested association mapping) project. This revealed strict co-inheritance of the unlinked *Rhg1* and  $NSF_{RAN07}$  alleles, demonstrating the functional necessity of *Rhg1* and  $NSF_{RAN07}$  co-occurrence.

### 5.3 Results

#### **Wild-type $\alpha$ -SNAP proteins are much less abundant while NSF is more abundant in *Rhg1*<sub>LowCopy</sub> soybeans**

We previously reported that the PI 88788-type *Rhg1*<sub>HC</sub> locus in soybean line “Fayette” drives a localized increase of resistance-type *Rhg1*  $\alpha$ -SNAP<sub>*Rhg1*HC</sub> protein to disrupt the developing SCN-induced syncytium (Bayless et al., 2016). We also reported that NSF levels increase when resistance-associated *Rhg1*  $\alpha$ -SNAP proteins are overexpressed (Bayless et al., 2016). However, for lines carrying low-copy Peking-type *Rhg1*<sub>LC</sub>, the cellular ratio of WT  $\alpha$ -SNAP to  $\alpha$ -SNAP<sub>*Rhg1*LC</sub> or NSF proteins was unknown. To investigate the relative abundances of WT and resistance-associated  $\alpha$ -SNAPs, in this study we performed immunoblots using the standard HG type test *Rhg1*<sub>HC</sub> and *Rhg1*<sub>LC</sub> soybean varieties and previously described anti- $\alpha$ -SNAP antibodies (Niblack et al., 2002; Bayless et al., 2016). We also studied NSF abundance in these samples using an antibody raised to a conserved NSF domain. As shown in Figure 1A, immunoblots from root tissue indicated that WT  $\alpha$ -SNAP abundance in all tested *Rhg1*<sub>LC</sub> lines (PI 548402/Peking, PI 90763, PI 437654, PI 89772) was dramatically reduced compared with the *Rhg1*<sub>HC</sub> lines (PI 88788, PI 209332, PI 548316). Probing of the same samples with antibodies that recognize  $\alpha$ -SNAP<sub>*Rhg1*LC</sub> or  $\alpha$ -SNAP<sub>*Rhg1*HC</sub> but not WT  $\alpha$ -SNAP confirmed that, between the *Rhg1*<sub>HC</sub> and *Rhg1*<sub>LC</sub> soybean varieties, there is a pronounced difference in the abundance of WT  $\alpha$ -SNAP relative to the abundance of *Rhg1*  $\alpha$ -SNAP (Fig 1A). As noted above, the reference soybean genome encodes five putative  $\alpha$ -SNAPs and the anti-WT- $\alpha$ -SNAP antibody was raised against the conserved C-terminus shared by all of those predicted WT  $\alpha$ -SNAP products, but not the resistance-associated *Rhg1*  $\alpha$ -SNAPs (Bayless et al., 2016). To interpret Fig. 1A, it is also useful to recall that the WT  $\alpha$ -SNAP genes include the *Rhg1* *Glyma.18G022500* gene, where one

of the *Rhg1<sub>HC</sub>* repeats encodes a WT  $\alpha$ -SNAP protein while all the other repeats encode a resistance-type *Rhg1*  $\alpha$ -SNAP protein (Cook et al., 2014; Lee et al., 2015). Yet low copy number *Rhg1<sub>LC</sub>* soybeans lack a *Glyma.18G022500* allele encoding WT  $\alpha$ -SNAP ( $\alpha$ -SNAP<sub>*Rhg1*</sub>WT) and encode only resistance-type  $\alpha$ -SNAP<sub>*Rhg1*</sub>LC protein at the Chromosome 18 locus (Cook et al., 2014; Lee et al., 2015). Additionally, in at least two previously studied *Rhg1<sub>LC</sub>* varieties, the Chromosome 11  $\alpha$ -SNAP gene *Glyma.11G234500* carries a mutated allele (Matsye et al., 2012; Lakhssassi et al., 2017).

NSF protein abundance in the *Rhg1<sub>LC</sub>* lines was increased compared with the *Rhg1<sub>HC</sub>* lines PI 88788 and PI 209332 (Fig. 1A, Fig. S1A). These differences in NSF expression, across two independent experiments, were quantified using densitometry with ImageJ (Fig 1B).

We then explored if WT  $\alpha$ -SNAP expression was similarly reduced in a more recent agriculturally utilized *Rhg1<sub>LC</sub>* soybean variety, “Forrest.” Immunoblots on both total leaf or root proteins from Williams82 (*Rhg1* single copy), Forrest (*Rhg1<sub>LC</sub>*) and Fayette (*Rhg1<sub>HC</sub>*), again revealed sharp decreases in total WT  $\alpha$ -SNAP abundance in the *Rhg1<sub>LC</sub>* source Forrest (Fig 1C). Altogether, a sharply reduced total abundance of WT  $\alpha$ -SNAPs was observed to be a shared trait of *Rhg1<sub>LC</sub>* soybean varieties but not *Rhg1<sub>HC</sub>* varieties. A likely hypothesis for this strikingly low abundance is the absence of a WT- $\alpha$ -SNAP-encoding allele at *Rhg1<sub>LC</sub>*, low or no product from the *Glyma.11G234500* ( $\alpha$ -SNAP<sub>Ch11</sub>) allele containing an intronic splice site mutation, and a relatively low contribution of protein from the other three putative  $\alpha$ -SNAP-encoding loci.

We also investigated if the native  $\alpha$ -SNAP<sub>*Rhg1*</sub>WT locus, if expressed, could contribute to total WT  $\alpha$ -SNAP protein abundance in *Rhg1<sub>LC</sub>* soybean lines. Cloning the native *Glyma.18G022500*  $\alpha$ -SNAP<sub>*Rhg1*</sub>WT locus from Williams 82 (Wm82), we generated transgenic Forrest (*Rhg1<sub>LC</sub>*) roots expressing native  $\alpha$ -SNAP<sub>*Rhg1*</sub>WT and assessed total WT  $\alpha$ -SNAP

abundance with immunoblots. Compared to empty vector controls, transgenic addition of the native Williams 82  $\alpha$ -SNAP<sub>Rhg1</sub>WT locus increased wild type  $\alpha$ -SNAP abundance in Forrest to levels similar to Williams 82 empty vector controls (Fig 1D).

**A unique NSF<sub>Ch07</sub> allele (RAN07) is present in Rhg1-containing NAM parents and HG type test type varieties**

Because *Rhg1*-resistance type  $\alpha$ -SNAPs ( $\alpha$ -SNAP<sub>Rhg1</sub>LC or  $\alpha$ -SNAP<sub>Rhg1</sub>HC) exhibited compromised binding to wild-type NSF and were toxic at high doses in *N. benthamiana* (Bayless et al., 2016), it was unclear how *Rhg1*LC lines could cope with the diminished WT  $\alpha$ -SNAP levels observed in Figure 1. NSF and  $\alpha$ -SNAP are essential housekeeping proteins in all eukaryotes and null mutations in either partner are lethal in animals, which typically encode only single copies of NSF or  $\alpha$ -SNAP (Littleton et al., 2001; Sanyal and Krishnan, 2001; Horsnell et al., 2002; Chae et al., 2004). Since soybean is a polyploid organism encoding multiple  $\alpha$ -SNAP and NSF loci, we examined whole genome sequence (WGS) data from multiple *Rhg1*-containing varieties for alterations in the other  $\alpha$ -SNAP or NSF loci. Briefly, we assembled reads for all  $\alpha$ -SNAP and NSF loci, and aligned them against the Williams 82 reference genome. In all  $\alpha$ -SNAP loci from *Rhg1*LC varieties, we detected no obvious polymorphisms other than the previously reported *Glyma.11G234500* ( $\alpha$ -SNAP<sub>Ch11</sub>) allele containing an intronic splice site mutation. However, among all examined *Rhg1*LC and *Rhg1*HC lines, a novel NSF<sub>Ch07</sub> allele was present containing five N-domain amino acid polymorphisms (R<sub>4</sub>Q, N<sub>21</sub>Y, S<sub>25</sub>N, <sup>^</sup><sub>116</sub>F, M<sub>181</sub>I) (Fig. 2A). Using cDNA from Forrest (*Rhg1*LC), we cloned and sequenced this unique NSF<sub>Ch07</sub> transcript and confirmed all 5 N-domain polymorphisms. Additionally, we designed two different PCR primer pairs at the N<sub>21</sub>Y and S<sub>25</sub>N polymorphisms and verified the presence of this unique NSF

$\text{Ch07}$  allele and absence of the wild-type  $\text{NSF}_{\text{Ch07}}$  allele in all HG type test lines, using agarose gel electrophoresis (S1B). Furthermore, using WGS data from the SoyNAM (Nested Association Mapping) project (Song et al., 2017), we determined that this unique  $\text{NSF}_{\text{Ch07}}$  allele was in every *Rhg1*-containing NAM parent, while SCN-susceptible NAM parents carried the WT  $\text{NSF}_{\text{Ch07}}$  allele (Table 1). We therefore named the protein from this *Rhg1*-associated allele of *Glyma.07G195900* "NSF<sub>RAN07</sub>" for "*Rhg1*-associated NSF from chromosome 07." In addition to  $\text{NSF}_{\text{RAN07}}$ , an allele of the chromosome 13 *Glyma.13g180100* gene encoding an  $\text{NSF}_{\text{Ch13}}$  V555I protein was found in some varieties, including SCN susceptible soybeans, but it was not present in all *Rhg1*<sub>LC</sub> or *Rhg1*<sub>HC</sub> lines (Table 1). Figure S2 shows the complete  $\text{NSF}_{\text{RAN07}}$  amino acid alignment to  $\text{NSF}_{\text{Ch07}}$  from the Williams 82 reference genome.

### **The $\text{NSF}_{\text{RAN07}}$ and *Rhg1* $\alpha$ -SNAP polymorphisms are both at the NSF/ $\alpha$ -SNAP binding interface**

The NSF/ $\alpha$ -SNAP interface consists of complementary electrostatic patches at the NSF N-domain and  $\alpha$ -SNAP C-terminus (Zhao et al., 2015; Zhao and Brunger, 2016). These binding patches are conserved in yeast, animals and plants, with the soybean NSF N-domain (N<sub>21</sub>, RR<sub>82-83</sub>, KK<sub>117-118</sub>) and  $\alpha$ -SNAP C-terminus (D<sub>208</sub>DEED<sub>243-246</sub>, EEDD<sub>284-287</sub>) corresponding to  $\text{NSF}_{\text{CHO}}$  (R<sub>10</sub>, RK<sub>67-68</sub>, KK<sub>104-105</sub>) and rat  $\alpha$ -SNAP (D<sub>217</sub>E<sub>249</sub>EE<sub>252-253</sub>, DEED<sub>290-293</sub>) respectively. Accordingly, inter-kingdom interactions between  $\alpha$ -SNAP and NSF have been reported both *in vitro* and for heterologous expression systems *in vivo*, including between soybean WT  $\alpha$ -SNAP and Chinese Hamster NSF ( $\text{NSF}_{\text{CHO}}$ ) (Griff et al., 1992; Bassham and Raikhel, 1999; Rancour et al., 2002; Bayless et al., 2016). To assess where the  $\text{NSF}_{\text{RAN07}}$  polymorphisms are positioned in the N-domain, we modeled  $\text{NSF}_{\text{RAN07}}$  to the  $\text{NSF}_{\text{CHO}}$  cryo-EM structure from Zhao and

colleagues (Zhao et al., 2015)(Fig. 2B). NSFs in many plants, including soybean, encode a variable length polyserine/glycine patch, starting at ~residue 6, hence modeling to NSF<sub>CHO</sub> began at residue 14. The NSF<sub>RAN07</sub> homology model to NSF<sub>CHO</sub> placed two of the NSF<sub>RAN07</sub> polymorphisms at two NSF<sub>CHO</sub> regions that bind  $\alpha$ -SNAP: N<sub>21</sub>Y at R<sub>10</sub> and S<sub>25</sub>N adjacent, and <sup>^</sup><sub>116</sub>F at RK<sub>114-115</sub>, respectively (Fig. 2B, C, S3A). While R<sub>4</sub>Q was omitted from the model, we examined R<sub>4</sub> frequency across 22 diverse eukaryotes (9 animals, 3 fungi, 10 plants) (Fig. 2D). In all but four model organisms, R<sub>4</sub> was present in the NSF of 18 of the 22 species, while *S. cerevisiae*, *Drosophila*, *C. elegans* and *Physcomitrella* carry an R and/or K at the adjacent residue #3 and/or #5. The final NSF<sub>RAN07</sub> polymorphism, M<sub>181</sub>I, was not located near the  $\alpha$ -SNAP binding patches and was not highly conserved among model organism NSFs. We additionally examined N-domain conservation in plant NSFs and determined that residues corresponding to N<sub>21</sub> and F<sub>115</sub> are present in a majority of plants and do not carry N<sub>21</sub>Y or the <sup>^</sup><sub>116</sub>F insertion (Fig. S3B). These results modeling just to NSF suggest that three of the five NSF<sub>RAN07</sub> N-domain polymorphisms are located in or adjacent to the NSF binding patches that interact with  $\alpha$ -SNAP.

The polymorphisms of both  $\alpha$ -SNAP<sub>Rhg1</sub>HC and  $\alpha$ -SNAP<sub>Rhg1</sub>LC, are located at conserved C-terminal residues that bind and stimulate NSF (Cook et al., 2014; Bayless et al., 2016). Multiple  $\alpha$ -SNAP proteins bound to a SNARE bundle recruit six NSF proteins to form a “20S supercomplex” (4X  $\alpha$ -SNAPs, 6X NSF, 3-4X SNAREs) and stimulate SNARE complex disassembly (Zhao et al., 2015). To further assess the proximity of the NSF<sub>RAN07</sub> N-domain polymorphisms to  $\alpha$ -SNAP C-terminal contacts, we identified and colored the complementary NSF and  $\alpha$ -SNAP binding residues, and then the NSF<sub>RAN07</sub> and Rhg1  $\alpha$ -SNAP polymorphisms, on the mammalian 20S cryo-EM structure (Fig. 3A, B, S4A, B). This confirmed that NSF<sub>RAN07</sub>

N<sub>21</sub>Y, S<sub>25</sub>N, <sup>116</sup>F are predicted to locate adjacent to NSF residues that bind  $\alpha$ -SNAP residues, including residues that contact WT  $\alpha$ -SNAP amino acid residues that are altered in  $\alpha$ -SNAP<sub>Rhg1</sub>HC and  $\alpha$ -SNAP<sub>Rhg1</sub>LC. R<sub>4</sub> on the NSF<sub>CHO</sub> structure was closely positioned to a D<sub>28</sub> side chain, present in soybean as D<sub>39</sub>(Fig. S4B). Altogether, the location and structural modeling of the NSF<sub>RAN07</sub> polymorphisms suggest that NSF<sub>RAN07</sub> modifies the normal NSF binding interface that maintains complementary binding contacts with  $\alpha$ -SNAP sites which are altered in Rhg1  $\alpha$ -SNAPs.

### **The NSF<sub>RAN07</sub> polymorphisms promote binding with *Rhg1* resistance-type $\alpha$ -SNAPs**

Since all *Rhg1*-containing HG type test and NAM lines contained NSF<sub>RAN07</sub>, and because  $\alpha$ -SNAP<sub>Rhg1</sub>HC and  $\alpha$ -SNAP<sub>Rhg1</sub>LC are polymorphic at C-terminal residues that bind and stimulate NSF, we investigated if the NSF<sub>RAN07</sub> polymorphisms impact binding with *Rhg1* resistance-type  $\alpha$ -SNAPs. Impacts of the NSF<sub>RAN07</sub> polymorphisms on binding to  $\alpha$ -SNAP<sub>Rhg1</sub>WT were also investigated. As in (Barnard et al., 1997) and (Bayless et al., 2016), we produced recombinant NSF<sub>RAN07</sub>, NSF<sub>Ch07</sub> and *Rhg1*  $\alpha$ -SNAP proteins for *in vitro* binding studies. As previously reported (Bayless et al., 2016), diminished NSF<sub>Ch07</sub> binding with  $\alpha$ -SNAP<sub>Rhg1</sub>HC and  $\alpha$ -SNAP<sub>Rhg1</sub>LC, compared to  $\alpha$ -SNAP<sub>Rhg1</sub>WT, was again observed (Fig. 3C). NSF<sub>RAN07</sub> binding to  $\alpha$ -SNAP<sub>Rhg1</sub>HC or  $\alpha$ -SNAP<sub>Rhg1</sub>LC, on the other hand, was more similar to  $\alpha$ -SNAP<sub>Rhg1</sub>WT binding and was increased ~30% relative to NSF<sub>Ch07</sub>. NSF<sub>RAN07</sub> and NSF<sub>Ch07</sub> binding was quantified using ImageJ densitometry across three independent experiments (Fig. 3D). Furthermore, to verify that NSF<sub>RAN07</sub>/ $\alpha$ -SNAP binding is dependent upon NSF-binding patches at the  $\alpha$ -SNAP C-terminus, we tested binding to an otherwise WT  $\alpha$ -SNAP lacking the final 10 C-terminal residues ( $\alpha$ -SNAP<sub>Rhg1</sub>WT<sub>1-279</sub>). Essentially, no binding of NSF<sub>Ch07</sub>WT or NSF<sub>RAN07</sub>

binding with  $\alpha$ -SNAP<sub>Rhg1</sub>WT<sub>1-279</sub> was observed, similar to the no  $\alpha$ -SNAP binding controls (Fig. S4C). Hence NSF<sub>RAN07</sub>/ $\alpha$ -SNAP binding requires the conserved NSF-binding contacts located at the  $\alpha$ -SNAP C-terminus. Combined, these binding assays suggest that NSF<sub>RAN07</sub> not only maintains normal binding to WT  $\alpha$ -SNAPs, but also at least partially accommodates the unusual C-terminal NSF-binding interface of *Rhg1* resistance-type  $\alpha$ -SNAPs.

### **The NSF<sub>RAN07</sub> polymorphisms guard against the cell death induced by *Rhg1*-resistance-type $\alpha$ -SNAP**

We previously observed that transient expression of either  $\alpha$ -SNAP<sub>Rhg1</sub>HC or  $\alpha$ -SNAP<sub>Rhg1</sub>LC in *N.benthamiana* leaves, via *Agrobacterium* infiltration, was cytotoxic and elicited a hyperaccumulation of the endogenous NSF protein (Bayless et al., 2016). Co-expression of WT- $\alpha$ -SNAP with the *Rhg1*  $\alpha$ -SNAP diminished this toxicity (Bayless et al., 2016). The penultimate leucine/isoleucine of  $\alpha$ -SNAP, which has been implicated in stimulation of NSF ATPase, was needed for rescue of this *N. benthamiana* cytotoxicity (Bayless et al., 2016). As such, we examined if soybean NSF co-expression might also alleviate the toxicity of *Rhg1* resistance-type  $\alpha$ -SNAPs in *N. benthamiana*. As in (Bayless et al., 2016), mixed *Agrobacterium* cultures containing 1 part WT  $\alpha$ -SNAP to 3 parts  $\alpha$ -SNAP<sub>Rhg1</sub>LC were used for cytotoxicity complementation assays. However, we noted that NSF<sub>RAN07</sub> and NSF<sub>Ch07</sub> were more effective than WT  $\alpha$ -SNAP at reducing *Rhg1*  $\alpha$ -SNAP cytotoxicity (Fig. S5A). We then decreased the proportion of NSF-delivering bacteria in the mixed *Agrobacterium* cultures down to 1 part to 9 or 14 parts  $\alpha$ -SNAP<sub>Rhg1</sub>LC-delivering bacteria. We observed that co-expressing soybean NSF<sub>Ch07</sub>, NSF<sub>Ch13</sub> or NSF<sub>RAN07</sub> reduced cell death caused by  $\alpha$ -SNAP<sub>Rhg1</sub>LC compared to empty vector controls (Fig. 4A), but that NSF<sub>RAN07</sub> co-expression consistently conferred greater

protection than either NSF<sub>Ch07</sub> or NSF<sub>Ch13</sub> (Fig. 4A). Infiltrated leaf patches had less death and/or slower death with NSF<sub>RAN07</sub>. Notably, both NSF<sub>RAN07</sub> and NSF<sub>Ch07</sub> were more effective than NSF<sub>Ch13</sub> at complementing cell death (Fig. 4A). NSF<sub>RAN07</sub> was even observed to confer at least partial protection out to a 1:19 mixture, again outperforming complementation by NSF<sub>Ch07</sub> (Fig. S5B). Complementation of  $\alpha$ -SNAP<sub>Rhg1</sub>HC-induced cell death with NSF<sub>RAN07</sub> vs. NSF<sub>Ch07</sub> produced similar results (Fig. S5C).

We also had previously observed elevated abundance of the endogenous *N. benthamiana* NSF upon expression of *Rhg1* resistance type- $\alpha$ -SNAPs, but, this does not prevent cell death (Bayless et al., 2016). However, it was unclear if immediate co-expression of NSF might lessen the cytotoxicity. Therefore, as in Fig. 4A, we agroinfiltrated mixed cultures of *N. benthamiana* NSF (NSF<sub>N.benth</sub>, 81% identity to NSF<sub>Ch07</sub>, see Fig. S6 for alignment) and  $\alpha$ -SNAP<sub>Rhg1</sub>LC, as well as EV, NSF<sub>Ch13</sub> and NSF<sub>RAN07</sub> as controls. As in Fig. 4A, NSF<sub>Ch13</sub> gave visible protection relative to an empty vector, while NSF<sub>RAN07</sub> co-expression gave strong protection (Fig. 4B). NSF<sub>N.benth</sub> co-expression, on the other hand, was similar to empty vector controls (Fig. 4B). As observed in (Bayless et al., 2016), expressing soybean NSFs or NSF<sub>N.benth</sub> with an empty vector at the same ratios used for complementation did not cause macroscopic phenotypes suggestive of stress (Fig. S5D).

Because no obvious complementation from co-expressing NSF<sub>N. benth</sub> was apparent, we examined physical binding with *Rhg1* resistance-type  $\alpha$ -SNAPs as in Fig. 3C, but using recombinant NSF<sub>N. benth</sub> protein. We observed that NSF<sub>N. benth</sub> readily bound  $\alpha$ -SNAP<sub>Rhg1</sub>WT, but binding to either *Rhg1* resistance-type  $\alpha$ -SNAP was much lower, only slightly over negative controls ( $\alpha$ -SNAP lacking the C-terminus or no- $\alpha$ -SNAP) (Fig. 4C). This suggests a biochemical explanation for why *Rhg1* resistance type  $\alpha$ -SNAPs - but not WT  $\alpha$ -SNAPs - provoke strong cell

death responses in *N. benthamiana*: the endogenous *N. benthamiana* NSF binds WT  $\alpha$ -SNAPs but not *Rhg1* resistance type  $\alpha$ -SNAPs. We therefore tested if cell-death caused by  $\alpha$ -SNAP<sub>*Rhg1*LC<sub>1-279</sub></sub>, which lacks the final 10 C-terminal residues and does not bind NSF<sub>RAN07</sub> or NSF<sub>Ch07</sub> *in vitro*, could be complemented by NSF<sub>RAN07</sub> or NSF<sub>Ch07</sub>. Neither NSF<sub>RAN07</sub> nor NSF<sub>Ch07</sub> prevented the cell death caused by  $\alpha$ -SNAP<sub>*Rhg1*LC<sub>1-279</sub></sub> whereas either complemented the cell death induced by full length  $\alpha$ -SNAP<sub>*Rhg1*LC</sub> (Fig. S5E). Finally, we explored how the penultimate  $\alpha$ -SNAP residue implicated in NSF-ATPase stimulation affected complementation by NSF<sub>RAN07</sub> or NSF<sub>Ch07</sub>, through complementation tests with  $\alpha$ -SNAP<sub>*Rhg1*LC I<sub>289A</sub></sub>. Complementation of  $\alpha$ -SNAP<sub>*Rhg1*LC I<sub>289A</sub></sub> was evident, but was less than that observed for  $\alpha$ -SNAP<sub>*Rhg1*LC</sub> (Fig 4D), suggesting that while NSF<sub>RAN07</sub> may bind *Rhg1* resistance type  $\alpha$ -SNAPs more effectively, ATPase-stimulation is likely an additional factor in relieving cytotoxicity. But overall, the findings of Figure 4 and related experiments extend the Figure 3 finding that NSF<sub>RAN07</sub> binds *Rhg1*  $\alpha$ -SNAPs better, by showing *in vivo* that NSF<sub>RAN07</sub> polymorphisms more effectively guard against the cell death induced by *Rhg1*  $\alpha$ -SNAPs.

**100% of the predicted *Rhg1*<sup>+</sup> *Glycine max* accessions in the USDA soybean collection, and 7% of the *rhg1*<sup>-</sup> accessions, contain the SoySNP50K NSF<sub>RAN07</sub> R4Q amino acid polymorphism**

NSF<sub>RAN07</sub> was present in all *Rhg1*-containing HG type and NAM lines, but we sought to test if this *Rhg1*/NSF<sub>RAN07</sub> association is universal rather than "frequent". We first sought to determine the approximate NSF<sub>RAN07</sub> allele frequency. In 2015, Song et al. reported genotyping the USDA soybean germplasm collection of ~20,000 accessions - collected from over 80 countries - using a 50,000 SNP DNA microarray chip (SoySNP50K iSelect BeadChip). The data

are available in a searchable SNP database at Soybase (Soybase.org/snps/) (Grant et al., 2010; Song et al., 2013, 2015). Using the Soybase genome browser, we found that a C/T SNP used on the SoySNP50K (ss715597431, Gm07:36,449,014) causes the NSF<sub>RAN07</sub> R<sub>4</sub>Q polymorphism. Analyzing all 19,645 USDA accessions for ss715597431, we estimated the NSF<sub>RAN07</sub> allele frequency in the USDA collection at 11.0% (2,165 +/+, 33 +/-). (Fig. 5A). While NSF in most model eukaryotes contains R<sub>4</sub>, it remained unclear whether Q<sub>4</sub> occurs in other plant NSFs. To determine if the NSF<sub>RAN07</sub> R<sub>4</sub>Q is unusual among plants, we examined R<sub>4</sub> conservation across plant NSF sequences available on Phytozome (Goodstein et al., 2012). Notably, Q<sub>4</sub> was not in the queried NSF predicted protein sequences for any other plant species (Fig. S7).

*Rhg1*-mediated SCN resistance is uncommon among soybean accessions and less than 5% of the USDA soybean collection carries a multi-copy *Rhg1* haplotype. Previously, Lee et al. identified SoySNP50K signatures for *Rhg1*<sub>HC</sub>, *Rhg1*<sub>LC</sub> and single copy (SCN-susceptible) haplotypes, and estimated that 705 *Rhg1*<sub>LC</sub> and 150 *Rhg1*<sub>HC</sub> accessions were in the USDA *Glycine max* collection (Lee et al., 2015). Using these 855 *Rhg1*-signature accessions, we determined a 100% incidence of the ss715597431 NSF<sub>RAN07</sub> signature for multi-copy *Rhg1*-signature *Glycine max* (Fig 5B).

If NSF<sub>RAN07</sub> is needed for the survival of *Rhg1*-containing soybean plants, then as we observed, all *Rhg1* accessions should carry NSF<sub>RAN07</sub>. As such, SNPs within the locus underlying *Rhg1* co-segregation should be maintained, while SNPs at neighboring loci, though tightly linked, would not be under stringent selection and hence should be less conserved. To narrow in on the *Rhg1* co-segregating locus within the interval, we examined amino acid changes within candidate loci adjacent to *RAN07* from *Rhg1*-carrying HG and NAM lines, between markers

ss715597415 and ss715597431. We observed that the  $NSF_{RAN07}$  SNPs, especially those causing the 5 N-domain polymorphisms, were 100% maintained across all *Rhg1*-containing varieties. On the other hand, SNPs causing amino acid changes within candidate loci adjacent to  $NSF_{RAN07}$ , were not 100% conserved across all *Rhg1*-containing varieties, unlike  $NSF_{RAN07}$  (Table S1). The predicted amino acid sequence of most candidate loci matches Wm82 (SCN-susceptible) sequence. Among candidate loci with amino acid substitutions, only *RAN07* has the same consistent amino acid changes across all examined *Rhg1*-containing germplasm. In addition to the observed biochemical and genetic complementation of *Rhg1*  $\alpha$ -SNAPs by  $NSF_{RAN07}$ , candidate gene allele frequency further implicate  $NSF_{RAN07}$  as the gene responsible for co-segregation with *Rhg1*.

**All *Rhg1*<sup>+</sup> F5-derived recombinant inbred lines (RILs) from NAM population crosses also carry  $NSF_{RAN07}$**

The above  $NSF_{RAN07}$  data from the USDA soybean germplasm collection are an indication of strong segregation distortion. However, recalling that Webb et al. (1995) reported that only 91 of 96 lines with a resistant parent marker type linked to *Rhg1* also had a resistant parent marker type near the  $NSF_{RAN07}$  QTL (Webb et al., 1995), we explored if lines with *Rhg1* strictly inherited  $NSF_{RAN07}$  in the progeny of more recent biparental crosses. From the Soybean Nested Associated Mapping (SoyNAM) project (Song et al., 2017), we examined genotypic data for populations of RILs developed from crosses of the IA3023 (SCN-susceptible) hub-parent to eight different soybean accessions carrying either *Rhg1*<sub>HC</sub> (seven accessions) or *Rhg1*<sub>LC</sub> (one accession). There were 122 to 139 RILs in each population and the segregation for  $NSF_{RAN07} : NSF_{Ch07WT}$  in soybean lines lacking *Rhg1* did not deviate from the null hypothesis of 1:1 segregation in six of

the eight populations. Across populations, there was a significant ( $\alpha=0.05$ ) deviation from a 1:1 segregation with a significantly greater number of RILs with  $NSF_{RAN07}$  than  $NSF_{Ch07WT}$ . The segregation distortion for  $NSF_{RAN07}$  was obvious among RILs that carried a resistance-associated  $Rhg1$  allele but, out of a total of 309  $Rhg1^+$  RILs, 8 appeared to have possibly inherited  $Rhg1_{HC}$  or  $Rhg1_{LC}$  but not  $NSF_{RAN07}$  while the remainder had  $NSF_{RAN07}$ . This was based upon the low-density SoySNP6K mapping data that did not include perfect genetic markers for  $Rhg1$  and  $NSF$ . We therefore genotyped for polymorphisms within  $Rhg1$  and  $NSF_{RAN07}$  genes, using primers that detect the  $Rhg1$  repeat junction and a WT  $NSF_{Ch07}$  vs.  $NSF_{RAN07}$  allele. All 8 re-examined RILs that inherited  $Rhg1_{HC}$  or  $Rhg1_{LC}$  also inherited the  $NSF_{RAN07}^{116F}$  and M<sub>181I</sub> mutations meaning that all 309 RILs that carried the resistance associated  $Rhg1$  also carried  $NSF_{RAN07}$  (Table S2). We analogously infer that in the Webb et al. 1995 study, the 5 lines that appeared to break the perfect co-inheritance between  $Rhg1_{HC}$  and  $NSF_{RAN07}$  were likely to have undergone a crossover between RFLP markers linked to either  $Rhg1$  or  $NSF$  (Webb et al., 1995). Taken together, the SoySNP50K and NAM data indicate that  $NSF_{RAN07}$  co-inheritance is a necessary balance that confers viability to soybeans that carry a multi-copy  $Rhg1$  haplotype.

### **The $\alpha$ -SNAP<sub>Ch11</sub>Intron Retention allele - a predicted SCN-resistance QTL - encodes an unstable protein**

A recent study implicated the locus carrying the intron-retention ( $\alpha$ -SNAP<sub>Ch11-IR</sub>) allele of  $\alpha$ -SNAP<sub>Ch11</sub> in SCN-resistance, but the responsible gene(s) within this QTL interval were not defined (Matsye et al., 2012; Lakhssassi et al., 2017). Importantly, it remained unclear if this  $\alpha$ -SNAP<sub>Ch11-IR</sub> transcript (Matsye et al., 2012) even produced a stable, albeit truncated, protein. We therefore cloned ORFs for both the wild-type  $\alpha$ -SNAP<sub>Ch11</sub> and the intron-retention ( $\alpha$ -SNAP<sub>Ch11-IR</sub>)

IR) version, added an N-terminal HA tag, and examined transient protein expression in *N. benthamiana*. Using anti-HA immunoblots, we observed that the N-HA- $\alpha$ -SNAP<sub>Ch11</sub> protein - but not the truncated N-HA- $\alpha$ -SNAP<sub>Ch11</sub>-IR protein - was readily detectable (Fig. 6A). To better assess the apparent instability of this truncated protein, we generated homology models of both  $\alpha$ -SNAP<sub>Ch11</sub>-IR and the WT  $\alpha$ -SNAP<sub>Ch11</sub> using the yeast  $\alpha$ -SNAP (sec17) crystal structure (Rice and Brunger, 1999). These models predict that the  $\alpha$ -SNAP<sub>Ch11</sub>IR protein terminates several residues into alpha-helix 12 (Fig 6B, S8A). We then used the non-tagged native WT  $\alpha$ -SNAP<sub>Ch11</sub> locus from Williams 82 to investigate if expression of the WT form of this chromosome 11 gene in an *Rhg1<sub>LC</sub>* genetic context can produce levels of  $\alpha$ -SNAP protein at all similar to those observed in Williams 82. As in Fig. 1D, we generated Forrest (*Rhg1<sub>LC</sub>*) transgenic roots but this time expressing the native Williams 82  $\alpha$ -SNAP<sub>Ch11</sub> locus, and determined the relative contributions of WT  $\alpha$ -SNAP<sub>Ch11</sub> to total WT  $\alpha$ -SNAP protein abundance with immunoblots (Fig. 6C). Compared to empty vector controls, transgenic addition of the native Williams 82  $\alpha$ -SNAP<sub>Ch11</sub> locus substantially boosted total WT  $\alpha$ -SNAP abundance in Forrest roots (Fig. 6C). We additionally observed the presence of a ~300 bp deletion in the promoter of the  $\alpha$ -SNAP<sub>Ch11</sub> IR allele in WGS data, and verified the presence of this deletion with PCR on genomic DNA from Cloud, Forrest and PI 89772 (Fig. S8B). Together, these results suggest that  $\alpha$ -SNAP<sub>Ch11</sub>WT protein can be expressed in viable *Rhg1<sub>LC</sub>* roots, but the intron-retention allele of  $\alpha$ -SNAP<sub>Ch11</sub> is undergoing (apparent) pseudogenization.

The  $\alpha$ -SNAP<sub>Ch11</sub>-IR allele with deficient  $\alpha$ -SNAP expression may have emerged randomly or it may confer some selective advantage, in particular regarding SCN resistance. For example, the  $\alpha$ -SNAP<sub>Ch11</sub> IR allele could contribute to SCN-resistance by reducing available levels of WT  $\alpha$ -SNAP proteins, shifting the balance of available  $\alpha$ -SNAP proteins toward the toxic Rhg1  $\alpha$ -

SNAP proteins. This could be particularly relevant in *Rhg1<sub>LC</sub>* soybean lines that typically carry only three copies of genes encoding  $\alpha$ -SNAP<sub>Rhg1<sub>LC</sub></sub> protein and correspondingly lower mRNA abundance for that product, in contrast to the nine- or ten-copy *Rhg1<sub>HC</sub>* lines (Cook et al., 2012; Cook et al., 2014). The complete loss of the  $\alpha$ -SNAP<sub>Rhg1<sub>WT</sub></sub> locus in *Rhg1<sub>LC</sub>* haplotypes could also contribute to the functional impacts of  $\alpha$ -SNAP<sub>Ch11<sub>WT</sub></sub> loss. We therefore used SoySNP50K data to analyze the frequency of the  $\alpha$ -SNAP<sub>Ch11</sub> *IR* allele in the whole USDA collection and in the 855 *Rhg1<sup>+</sup>* *Glycine max* accessions noted above. In the USDA collection, the  $\alpha$ -SNAP<sub>Ch11</sub> *IR*-associated ss715610416 genotype was present in 5.6% of accessions (Fig. 6D). Perhaps surprisingly, we observed the  $\alpha$ -SNAP<sub>Ch11</sub> *IR*-associated ss715610416 genotype in roughly half (55.9%) of the *Rhg1<sub>LC</sub>* soybean lines and in about a third (34.7%) of the *Rhg1<sub>HC</sub>* lines (Fig. 6E). Only a subtle positive impact on SCN resistance was reported for the broader QTL locus carrying the  $\alpha$ -SNAP<sub>Ch11</sub> *IR* allele (Lakhssassi et al., 2017), but SCN resistance data that quantitatively reflect field-based yield improvement under SCN pressure are difficult to obtain. Not all commercial *Rhg1<sup>+</sup>* soybean lines carry the  $\alpha$ -SNAP<sub>Ch11</sub> *IR* allele, and its use or exclusion may translate to subtle but economically useful shifts in SCN resistance, or in the HG type specificity of that resistance, or in soybean yield potential in the presence of *Rhg1*, *Rhg4* and other SCN resistance QTLs.

## 5.4 Discussion

Across eukaryotes, NSF and  $\alpha$ -SNAP interface through conserved electrostatic contacts to disassemble SNARE complexes, thereby maintaining cellular vesicle fusion (Jahn and Scheller, 2006; Zhao and Brunger, 2015). This study indicates that *Rhg1*-mediated SCN resistance results not only from an unusual change in the *Rhg1*  $\alpha$ -SNAP sequence and from  $\alpha$ -SNAP accumulation in syncytium cells, as previously published, but also from changes in other housekeeping  $\alpha$ -SNAP and NSF genes whose products comprise the SNARE-recycling machinery. This study also suggests that the two distinct resistance-conferring *Rhg1* haplotypes employ similar yet distinct strategies to combat SCN. They decrease WT  $\alpha$ -SNAP availability through disparate means, via significant copy number expansion and/or through loss of wild-type  $\alpha$ -SNAP loci. This implicates WT  $\alpha$ -SNAPs as key factors in SCN syncytium formation. We also found that presence of the unusual *Rhg1*  $\alpha$ -SNAP proteins requires co-presence of a novel NSF protein for plant viability. This explains a well-documented segregation distortion occurring between *Rhg1* and a chromosome 7 region (Webb et al., 1995; Kopisch-Obuch and Diers, 2006; Vuong et al., 2015), but perhaps more importantly, the combined set of findings in the present study and other recent work on *Rhg1* offer a molecular framework in which to understand the interactions of multiple QTLs associated with SCN resistance: Many of these loci modify the host vesicle fusion SNARE recycling machinery as a means of controlling SCN infection success.

An understanding of the necessity of NSF<sub>RAN07</sub> to balance *Rhg1* germplasm should become a central consideration in any planned transgenic addition of *Rhg1* into SCN-susceptible soybeans. Beyond soybean, the present findings suggest strategies to engineer *Rhg1*-like resistance into other cyst nematode-susceptible crop species, through modulation of WT  $\alpha$ -SNAP

abundance, introduction of sequence-edited  $\alpha$ -SNAP alleles and/or introduction of a compatible NSF.

It is biologically fascinating that complementary  $\alpha$ -SNAP and NSF polymorphisms, located at the conserved binding interfaces of both members of the core SNARE recycling machinery, were apparently selected due to disease pressure from SCN. This underscores this pathway's importance during the pathogen-host interaction. The previous finding that *Rhg1* resistance-type  $\alpha$ -SNAPs are impaired in normal NSF-interactions (Bayless et al., 2016) is supported by the present finding that a unique NSF allele - NSF<sub>RAN07</sub> - is a requisite balance for *Rhg1* resistance  $\alpha$ -SNAPs. While (Bayless et al., 2016) proposed the functional redundancy of multiple WT  $\alpha$ -SNAP loci (available due to polyploidy) as the balance that allows the viability of *Rhg1*-containing lines, this model must be modified with the observation that *Rhg1*-containing lines that lack NSF<sub>RAN07</sub> are not viable. Presence of WT  $\alpha$ -SNAPs may still, in the presence of NSF<sub>RAN07</sub>, contribute to the viability and normal soybean yield of lines carrying the PI 88788 source of *Rhg1* (*Rhg1*<sub>HC</sub>), but they are not sufficient to do so in the absence of NSF<sub>RAN07</sub>. However, the observation that *Rhg1*<sub>LC</sub> varieties exhibit sharply reduced WT  $\alpha$ -SNAP expression further supports the idea that WT  $\alpha$ -SNAP levels and the WT  $\alpha$ -SNAP : *Rhg1*  $\alpha$ -SNAP ratio also are important components of successful *Rhg1*-mediated SCN resistance.

As noted above, the present findings about NSF<sub>RAN07</sub> provide a mechanistic explanation for soybean breeder reports that describe segregation distortion between *Rhg1* and the chromosome 7 genetic interval that encodes NSF<sub>RAN07</sub> in resistant plants (Webb et al., 1995; Kopisch-Obuch and Diers, 2006; Vuong et al., 2015). An observation that remains unexplained, however, is why transgenic expression of  $\alpha$ -SNAP<sub>*Rhg1*HC</sub> or  $\alpha$ -SNAP<sub>*Rhg1*LC</sub> protein, in *Agrobacterium rhizogenes*-transformed root systems of SCN-susceptible Williams82 (which

lacks *NSF<sub>RAN07</sub>*), elicited no apparent sensitivities such as cytotoxicity or endogenous NSF expression increases (Cook et al., 2012; Bayless et al., 2016). These sensitivities were observed with *N. benthamiana* expressing *Rhg1*  $\alpha$ -SNAP (Bayless et al., 2016). Notably, co-expression of *NSF<sub>N.benth</sub>* did not relieve the cell death in *N. benthamiana* leaves caused by *Rhg1*  $\alpha$ -SNAP, while WT soybean *NSF<sub>Ch07</sub>* did, albeit not as well as *NSF<sub>RAN07</sub>*. Consistent with this, recombinant *NSF<sub>N.benth</sub>* essentially could not bind with *Rhg1* resistance-type  $\alpha$ -SNAPs *in vitro* while soybean WT *NSF<sub>Ch07</sub>* binding was detectable. This may explain why soybean root cells do exhibit some tolerance of *Rhg1*  $\alpha$ -SNAP expression even in the absence of *NSF<sub>RAN07</sub>*. Nevertheless, the observation that all soybeans in the USDA collection that bear the signature of resistance-conferring *Rhg1* alleles also contain *NSF<sub>RAN07</sub>*, coupled with the universal co-presence of the *NSF<sub>RAN07</sub>* allele when *Rhg1* is present in the segregating progeny of NAM crosses, provides compelling evidence that at the organismal level, *NSF<sub>RAN07</sub>* is essential for the viability at some stage of growth of all *Rhg1*-containing germplasm.

While divergent evolution has created differences in copy number and  $\alpha$ -SNAP alleles between the *Rhg1<sub>HC</sub>* and *Rhg1<sub>LC</sub>* haplotypes, both accomplish similar things through alteration of the ratio of WT to *Rhg1* resistance-type  $\alpha$ -SNAPs. It remains unclear, however, how replacing or reducing the levels of WT  $\alpha$ -SNAPs is beneficial to SCN resistance. Overexpression of resistance-associated *Rhg1*  $\alpha$ -SNAPs did disrupt both exocytosis of a sec-GFP marker and localization of the Syp61 trans-Golgi network marker in *N. benthamiana* leaves (Bayless et al., 2016). The level of resistance-associated *Rhg1*  $\alpha$ -SNAPs relative to WT  $\alpha$ -SNAPs increases in syncytium cells of infected soybean roots (Bayless et al., 2016). This suggested a model in which elevated levels of the dysfunctional *Rhg1*  $\alpha$ -SNAP poison the syncytium, thereby disrupting the biotrophic interface between nematode and plant that cyst nematodes rely on for growth and

reproduction (Bayless et al., 2016). The present findings add to what was already known or inferred about loss of some WT  $\alpha$ -SNAPs in Peking-type *Rhg1<sub>LC</sub>* soybean lines (Matsye et al., 2012; Cook et al., 2014; Lee et al., 2015; Lakhssassi et al., 2017), providing evidence that in addition to the unusual *Rhg1*  $\alpha$ -SNAP proteins, altered levels of WT  $\alpha$ -SNAPs can also contribute to SCN resistance.

*Rhg1<sub>LC</sub>* and *Rhg4* contribute together to the SCN resistance of *Rhg1<sub>LC</sub>* soybean lines (Liu et al., 2012; Mitchum, 2016), and it remains unclear why *Rhg1<sub>LC</sub>* confers only partial SCN resistance in *Rhg1<sub>LC</sub>/Rhg1<sub>LC</sub> rhg4<sup>-</sup>/rhg4<sup>-</sup>* soybeans (Brucker et al., 2005; Liu et al., 2012; Yu et al., 2016). The present study characterized  $\alpha$ -SNAP and NSF features of both *Rhg1<sub>LC</sub>* and *Rhg1<sub>HC</sub>* soybean lines. Whether or not the *Rhg4* product directly impacts *Rhg1*-associated  $\alpha$ -SNAP/NSF/SNARE interactions, the published evidence suggests that *Rhg1<sub>HC</sub>* soybean lines are substantially more effective at conferring SCN resistance against HG type 0 SCN populations than *Rhg1<sub>LC</sub> rhg4<sup>-</sup>* soybeans .

Discovery of the need for NSF<sub>RAN07</sub> in *Rhg1*-containing soybeans may reveal a protective mechanism that reduces the toxicity of *Rhg1*  $\alpha$ -SNAPs in some cell types/conditions by facilitating participation of *Rhg1*  $\alpha$ -SNAPs in productive 20S complexes that disassemble SNARE bundles, while the toxicity of *Rhg1*  $\alpha$ -SNAPs remains predominant in syncytium cells. However, other mechanistic hypotheses are viable. Unfortunately, the tools to distinguish between the above two hypotheses are not yet available. Additional future studies could examine the dynamics of NSF<sub>RAN07</sub> abundance and function over time in developing SCN syncytia. For example, increased NSF levels were detected in syncytia in *Rhg1<sub>HC</sub>* varieties, and we had associated this with  $\alpha$ -SNAP deficiency (Bayless et al., 2016), but whether it is NSF<sub>RAN07</sub> or NSF<sub>Ch13</sub> that increases is of obvious interest and might suggest whether  $\alpha$ -SNAP and NSF

functionality is being promoted or disrupted by the host. We did observe that  $NSF_{RAN07}$  apparently can work with WT  $\alpha$ -SNAPs, or at least is not toxic in the way that resistance-associated  $Rhg1$   $\alpha$ -SNAPs can be toxic. Expression of  $NSF_{RAN07}$  in *N. benthamiana* caused no macroscopically detectable leaf phenotypes and it is expressed in  $Rhg1_{HC}$  soybeans that also express high levels of WT  $\alpha$ -SNAPs. The random 1:2:1 segregation of the alleles encoding  $NSF_{Ch07}^{WT}$  and  $NSF_{RAN07}$  in soybean progeny that lack  $Rhg1$ , and the presence of  $NSF_{RAN07}$  in over 1300 USDA soybean accessions that lack  $Rhg1$ , also suggests that presence of  $NSF_{RAN07}$  on its own is relatively benign.

The amassing evidence for the importance of  $\alpha$ -SNAP/NSF/SNARE interactions in SCN-soybean interactions also suggests these proteins as important targets for cyst nematode effectors. Preliminary evidence for one such effector is already in place (Bekal et al., 2015). A major issue for global soybean production at present is the gradual evolution of many SCN populations toward a capacity of some nematode individuals to overcome the widely used  $Rhg1_{HC}$  SCN resistance. Future work to discover and understand relevant nematode effectors in these SCN populations, and a means of re-establishing resistance against such nematodes, may benefit from assays that directly test for effectors that impact the soybean  $\alpha$ -SNAP and NSF protein variants characterized in the present study.

## 5.5 Materials & Methods

### Recombinant Protein Production

Vectors encoding recombinant  $\alpha$ -SNAP<sub>RhgI</sub>HC,  $\alpha$ -SNAP<sub>RhgI</sub>LC,  $\alpha$ -SNAP<sub>RhgI</sub>WT,  $\alpha$ -SNAP<sub>RhgI</sub>WT<sub>1-285</sub> and the WT alleles of NSF *Glyma.07G195900* (NSF<sub>Ch07</sub>) and *Glyma.13G180100* (NSF<sub>Ch13</sub>) were generated in Bayless et al., 2016. The open reading frames (ORFs) encoding the soybean NSF<sub>RAN07</sub> allele of *Glyma.07G195900* or *N.benthamiana* NSF were cloned into the expression vector pRham N-His-SUMO Kan according to manufacturer instructions (Lucigen). Recombinant  $\alpha$ -SNAP and NSF proteins were also produced and purified as in Bayless et al. 2016. All expression constructs were chemically transformed into the expression strain “E. cloni 10G” (Lucigen), grown to OD<sub>600</sub> ~0.60-0.70, and induced with 0.2% L-Rhamnose (Sigma) for either 8 hr at 37°C or overnight at 28°C. Soluble, native recombinant His-SUMO- $\alpha$ -SNAPs or His-SUMO-NSF proteins were purified with PerfectPro Ni-NTA resin (5 PRIME), with similar procedures as described in (Bayless et al., 2016) and eluted with imidazole, though no subsequent gel filtration steps were performed. Following the elution of the His-SUMO–fusion proteins, overnight dialysis was performed at 4 °C in 20 mM Tris (pH 8.0), 150 mM NaCl, 10% (vol/vol) glycerol, and 1.5 mM Tris (2-carboxyethyl)-phosphine. The His-SUMO affinity/solubility tags were cleaved from  $\alpha$ -SNAP or NSF using 1 or 2 units of SUMO Express protease (Lucigen) and separated by rebinding of the tag with Ni-NTA resin and collecting the recombinant protein from the flowthrough. Recombinant protein purity was assessed by Coomassie blue staining and quantified via a spectrophotometer.

### *In vitro* NSF- $\alpha$ -SNAP Binding Assays

*In vitro* NSF binding assays were performed essentially as described in (Barnard et al., 1996; Bayless et al., 2016). Briefly, 20  $\mu\text{g}$  of each respective recombinant  $\alpha$ -SNAP protein was added to the bottom of a 1.5-mL polypropylene tube and incubated at 25°C for 20 min. Unbound  $\alpha$ -SNAP proteins were then washed by adding  $\alpha$ -SNAP wash buffer [25 mM Tris, pH 7.4, 50 mM KCl, 1 mM DTT, 0.4 mg/mL bovine serum albumin (BSA)]. After removal of wash buffer, 20  $\mu\text{g}$  of recombinant NSF (1  $\mu\text{g}/\mu\text{L}$  in NSF binding buffer), was then immediately added and incubated on ice for 10 min. The solution was then removed and samples were immediately washed 2X with NBB to remove any unbound NSF. Samples were then boiled in 1X SDS loading buffer and separated on a 10% Bis-Tris SDS-PAGE, and silver-stained using the ProteoSilver Kit (Sigma-Aldrich), according to the manufacturer directions. The percentage of NSF bound by  $\alpha$ -SNAP was then calculated using densitometric analysis with ImageJ.

### **Antibody Production and Validation**

Affinity-purified polyclonal rabbit antibodies raised against  $\alpha$ -SNAP<sub>Rhg1</sub>HC,  $\alpha$ -SNAP<sub>Rhg1</sub>LC and wild-type  $\alpha$ -SNAPs were previously generated and validated using recombinant proteins in Bayless 2016. The epitopes for these custom antibodies are the final six or seven C-terminal  $\alpha$ -SNAP residues: “EEDDLT,” “EQHEAIT,” or “EEYEVIT” for wild-type, high-, or low-copy  $\alpha$ -SNAPs, respectively. For NSF, a synthetic peptide, “ETEKNVRDLFADAEQDQRTRGDESD,” corresponding to residues 300 to 324 of *Glyma.07G195900* was used. This NSF antibody was previously shown to be cross-reactive with the *N.benthamiana*-encoded NSF.

### **Immunoblotting**

Tissue preparation and immunoblots were performed essentially as in (Song et al., 2015; Bayless et al., 2016). Soybean roots or *N. benthamiana* leaf tissues were flash-frozen in N<sub>2</sub>(L), massed, and homogenized in a PowerLyzer 24 (MO BIO) for three cycles of 15 seconds, with flash-freezing in-between each cycle. Protein extraction buffer [50 mM Tris·HCl (pH 7.5), 150 mM NaCl, 5 mM EDTA, 0.2% Triton X-100, 10% (vol/vol) glycerol, 1/100 Sigma protease inhibitor cocktail] was then added at a 3:1 volume to mass ratio and samples were centrifuged and stored on ice. In noted experiments, Bradford assays were performed on each sample, and equal OD amounts of total protein were loaded in each sample lane for SDS/PAGE. Immunoblots for either Rhg1  $\alpha$ -SNAP were incubated overnight at 4 °C in 5% (wt/vol) nonfat dry milk TBS-T (50 mM Tris, 150 mM NaCl, 0.05% Tween 20) at 1:1,000. NSF immunoblots were performed similarly, except incubations were for 1 h at room temperature. Secondary horseradish peroxidase-conjugated goat anti-rabbit IgG was added at 1:10,000 and incubated for 1 h at room temperature on a platform shaker, followed by four washes with TBS-T. Chemiluminescence detection was performed with SuperSignal West Pico or Dura chemiluminescent substrate (Thermo Scientific) and developed using a ChemiDoc MP chemiluminescent imager (Bio-Rad).

### **Transgenic Soybean Root Generation**

Binary expression constructs were transformed into *Agrobacterium rhizogenes* strain, “Arqua1”. Transgenic soybean roots were produced as described in (Cook et al., 2012).

**Transient *Agrobacterium* Expression in *Nicotiana benthamiana*.** *Agrobacterium tumefaciens* strain GV3101 was used for transient protein expression of all constructs via syringe-infiltration at OD<sub>600</sub> 0.60 for NSF constructs or OD<sub>600</sub> 0.80 for  $\alpha$ -SNAP constructs into young leaves of ~4-

wk-old *N. benthamiana* plants. GV3101 cultures were grown overnight at 28°C in 25 µg/mL kanamycin and rifampicin and induced for ~3.5 h in 10 mM Mes (pH 5.60), 10 mM MgCl<sub>2</sub>, and 100µM acetosyringone prior to leaf infiltration. *N. benthamiana* plants were grown in a Percival set at 25 °C with a photoperiod of 16 h light at 100 µE·m<sup>-2</sup>·s<sup>-1</sup> and 8 h dark. For α-SNAP complementation assays, GV3101 cultures were well-mixed with one volume of an empty vector control, or of the respective NSF construct immediately before co-infiltration. NSF<sub>RAN07</sub> or the *N. benthamiana* NSF were PCR amplified from a root cDNA library of *Rhg1*<sub>LC</sub> variety, “Forrest”. or a *N.benthamiana* leaf cDNA library using KAPA HiFi polymerase, respectively. Expression cassettes for NSF<sub>*N.benthamiana*</sub>, NSF<sub>Ch13</sub>, NSF<sub>Ch07</sub> and NSF<sub>RAN07</sub> ORFs were directly assembled into a pBluescript vector containing the soybean ubiquitin (GmUbi) promoter and NOS terminator using Gibson assembly. The NSF expression cassettes were then digested with the restriction enzymes NotI-SalI and ligated with T4 DNA ligase into the previously described binary vector, pSM101-linker, which was cut with PspOMI-SalI restriction sites. The ORF encoding the α-SNAP<sub>Ch11</sub> Intron-Retention (IR) allele was amplified with Kapa HiFi from a root cDNA library of *Rhg1*<sub>LC</sub> variety “Forrest” while the ORF encoding WT α-SNAP<sub>Ch11</sub> was previously generated in (Bayless et al., 2016). Both α-SNAP<sub>Ch11</sub> and α-SNAP<sub>Ch11</sub>IR were Gibson assembled into a pBluescript vector containing a GmUbi-N-HA tag and NOS terminator, cut with PstI-XbaI and ligated into the binary vector, pSM101, cut with the same restriction pair. An 11.14 kb native genomic region encoding α-SNAP<sub>*Rhg1*</sub>WT was amplified with Kapa HiFi from a previously described fosmid subclone (Fosmid 19) with AvrII-SbfI restriction ends, and then digested and ligated into the binary vector, pSM101, cut with XbaI-PstI. A 6.85 kb native locus encoding α-SNAP<sub>Ch11</sub> was amplified from gDNA of Williams82 into two fragments (3.25 kb and 3.60 kb fragments) and Gibson assembled into pSM101 vector cut with BamHI-PstI.

### **Segregating NAM Crosses**

These segregating soybean crosses and mapping were developed and reported in (Song et al., 2017)

### **Protein Structure Modeling and Sequence Logo**

NSF<sub>RAN07</sub>,  $\alpha$ -SNAP<sub>Ch11</sub> and  $\alpha$ -SNAP<sub>Ch11IR</sub> structural homology models were generated using SWISS-MODEL and output PDB files viewed and labeled using PyMol. NSF<sub>RAN07</sub> was modeled to NSF<sub>CHO</sub> (Chinese hamster ovary) (PDB 3j97.1) cryo-EM structure from Zhao *et al* (Brunger group). 20S supercomplex modeling also generated using PDB 3j97, with  $\alpha$ -SNAPs and SNAREs of *Rattus norvegicus* origin (Zhao et al., 2015).  $\alpha$ -SNAP<sub>Ch11</sub> and  $\alpha$ -SNAP<sub>Ch11IR</sub> were modeled to sec17 (yeast  $\alpha$ -SNAP) crystal structure 1QQE donated courtesy of Rice *et al* (Brunger group)(Rice and Brunger, 1999).

The R<sub>4</sub>Q NSF amino acid consensus logo was generated from the first 10 NSF amino acids of the model eukaryotic organisms given in Fig. 2D. using WebLogo (Crooks et al., 2004)

### **DNA Sequence and SNP Analysis**

Whole-genome sequencing data of 12 soybean varieties was obtained from previously published studies (Cook et al., 2014; Song et al., 2017). Illumina sequencing reads were aligned to the Williams 82 reference genome (Wm82.a2.v1) using BWA (version 0.7.12)(Li and Durbin, 2009). Reads were initially mapped using the default settings of the *aln* command with the subsequent pairings performed with the *sampe* command. Alignments were next processed using the program Picard (version 2.9.0) to add read group information (AddOrReplaceReadGroups), mark PCR duplicates (MarkDuplicates, and merge alignments from separate sequencing runs

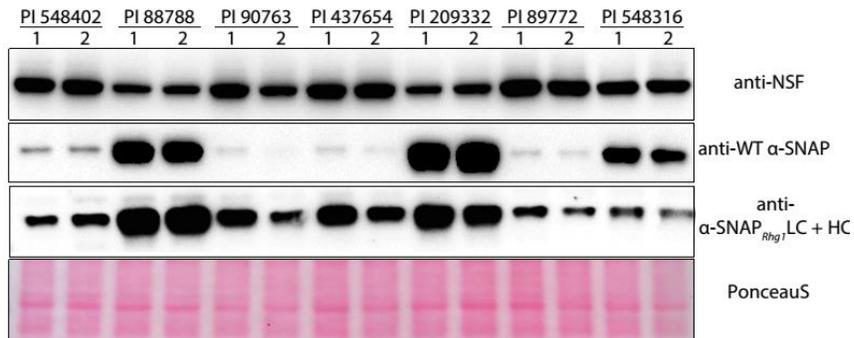
(MergeSamFiles). The processed .bam files were then converted to vcf format using a combination of samtools (version 0.1.19) and bcftools (version 0.1.19). Finally, consensus sequences were generated from these .vcf files using the FastaAlternateReferenceMaker tool within GATK (version 3.7.0)(DePristo et al., 2011).

## 5.6 Acknowledgments

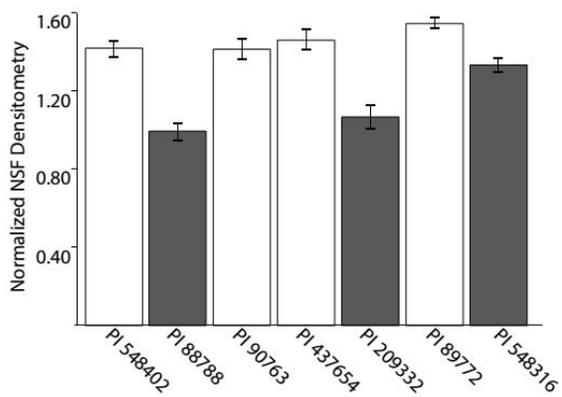
Thanks to Kaela Amundson for growing and taking care of all the *N. benthamiana* plants. This work was funded primarily by USDA-NIFA-AFRI award 2014-67013-21775 to A.F.B., and also by the United Soybean Board. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. (DGE-1256259) to A.M.B.

## 5.7. Figures

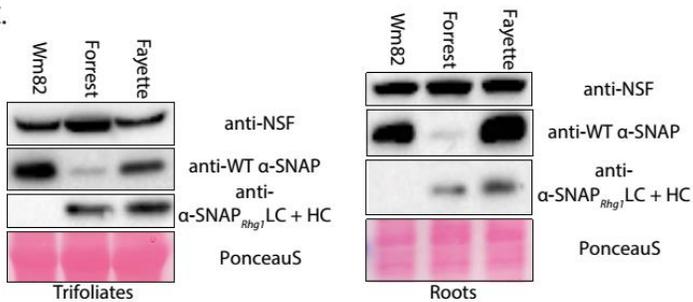
1A.



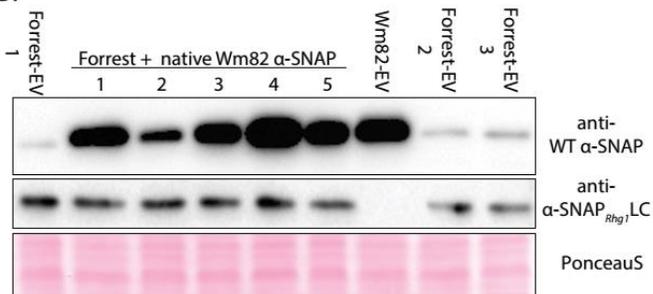
1B.



1C.

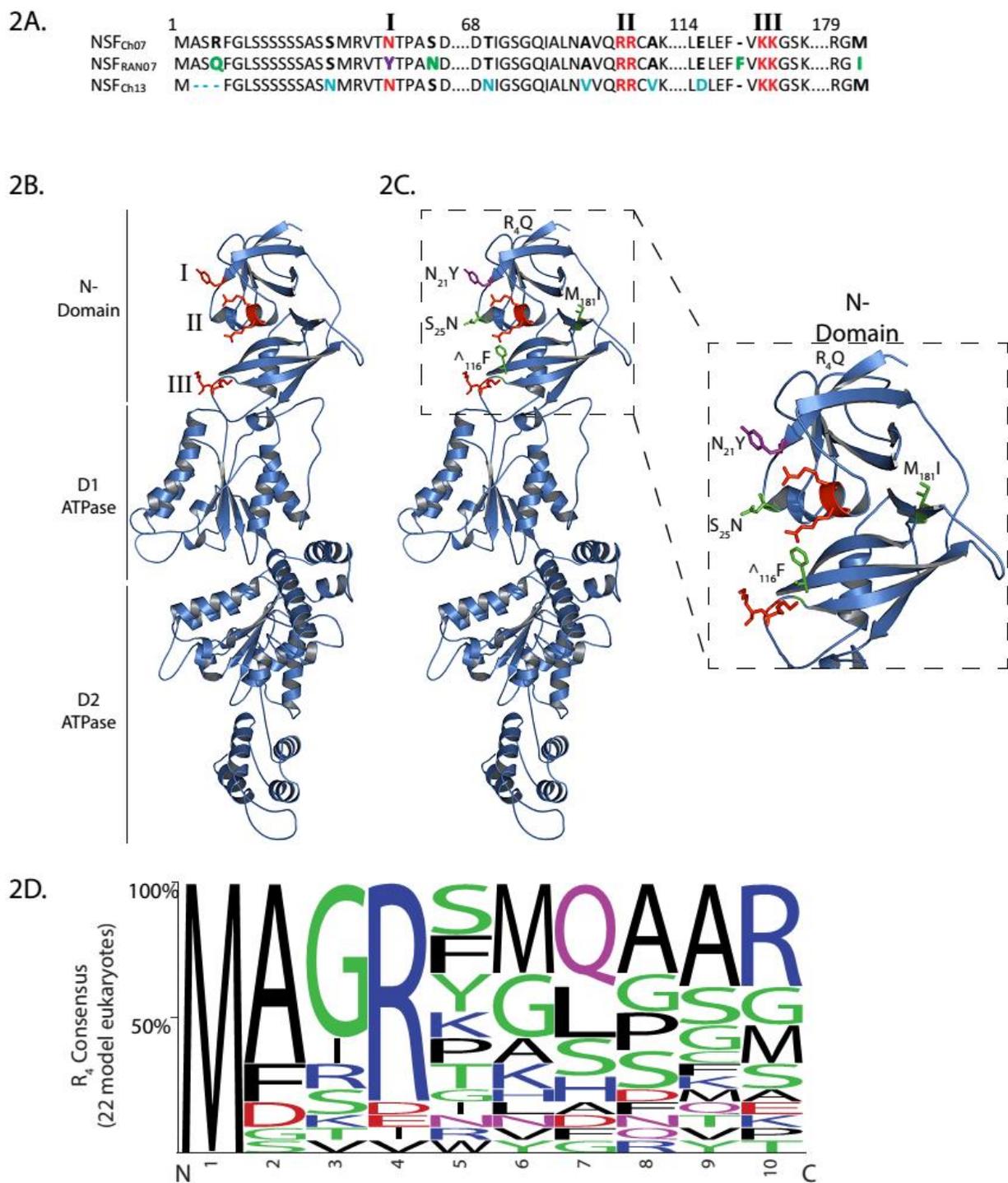


1D.



**Fig. 1.** Wild-type  $\alpha$ -SNAP expression is reduced in *RhgI*<sub>LowCopy</sub> soybeans;  $\alpha$ -SNAP<sub>*RhgI*WT</sub> is the predominant soybean  $\alpha$ -SNAP

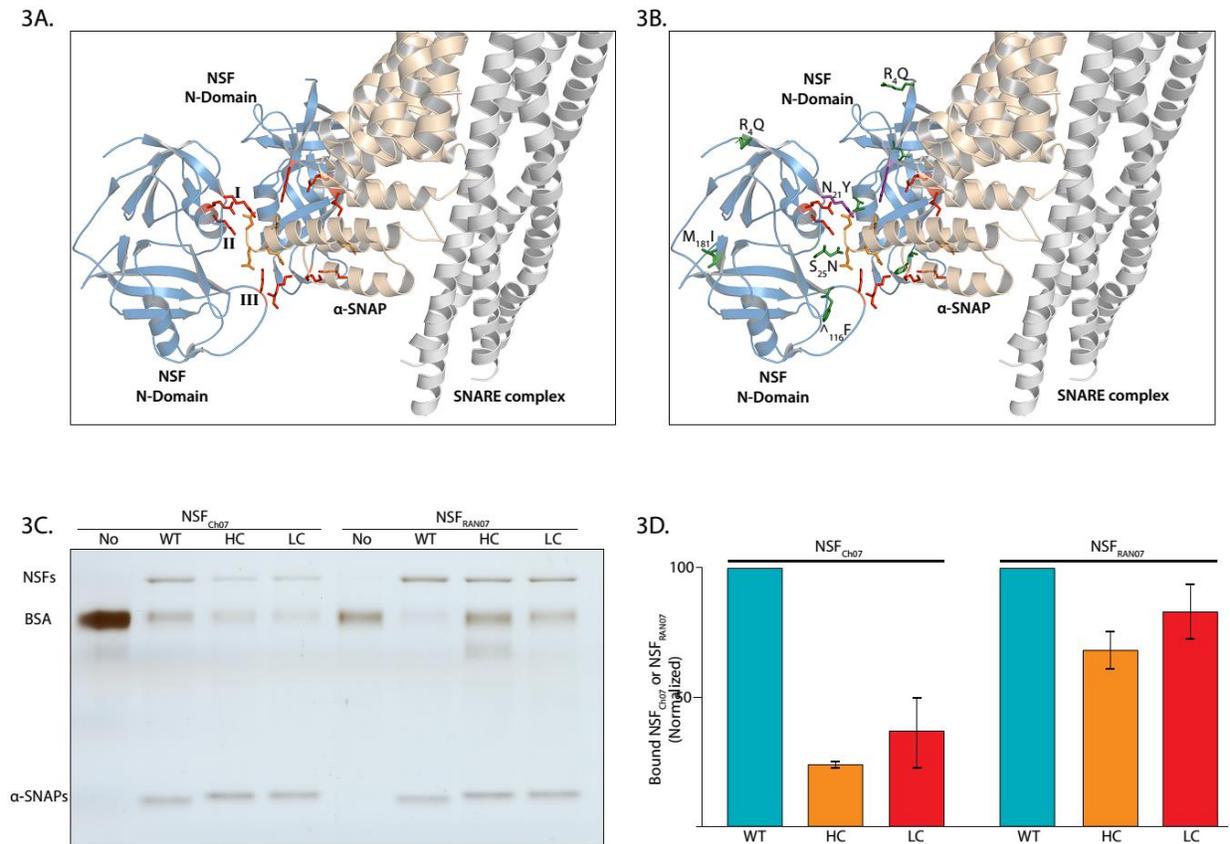
(A). Immunoblot of wild-type  $\alpha$ -SNAPs, *RhgI* resistance-type  $\alpha$ -SNAPs and NSF in HG type test soybean roots. *RhgI*<sub>LC</sub> varieties: PI 548402 (Peking), PI 89772, PI 437654, PI 90763; *RhgI*<sub>HC</sub> varieties: PI 88788, PI 209332, PI 548316(7 copy). PonceauS staining shows total protein loaded per lane. (B). Densitometry indicating total NSF expression in HG type test lines. (C). Like A, but immunoblots on trifoliolate leaves or roots of Williams 82 (Wm82) and modern *RhgI*<sub>LC</sub> and *RhgI*<sub>HC</sub> varieties Forrest and Fayette. (D). Immunoblots for total WT  $\alpha$ -SNAPs and  $\alpha$ -SNAP<sub>*RhgI*LC</sub> in “Forrest” (*RhgI*<sub>LC</sub>) transgenic roots transformed with an empty vector (EV) or the native Williams 82  $\alpha$ -SNAP<sub>*RhgI*WT</sub> locus, or in Williams 82 roots transformed with empty vector.



**Fig. 2.** *Rhg1*-containing lines carry a NSF<sub>Ch07</sub> allele (*RAN07*) with N-domain polymorphisms

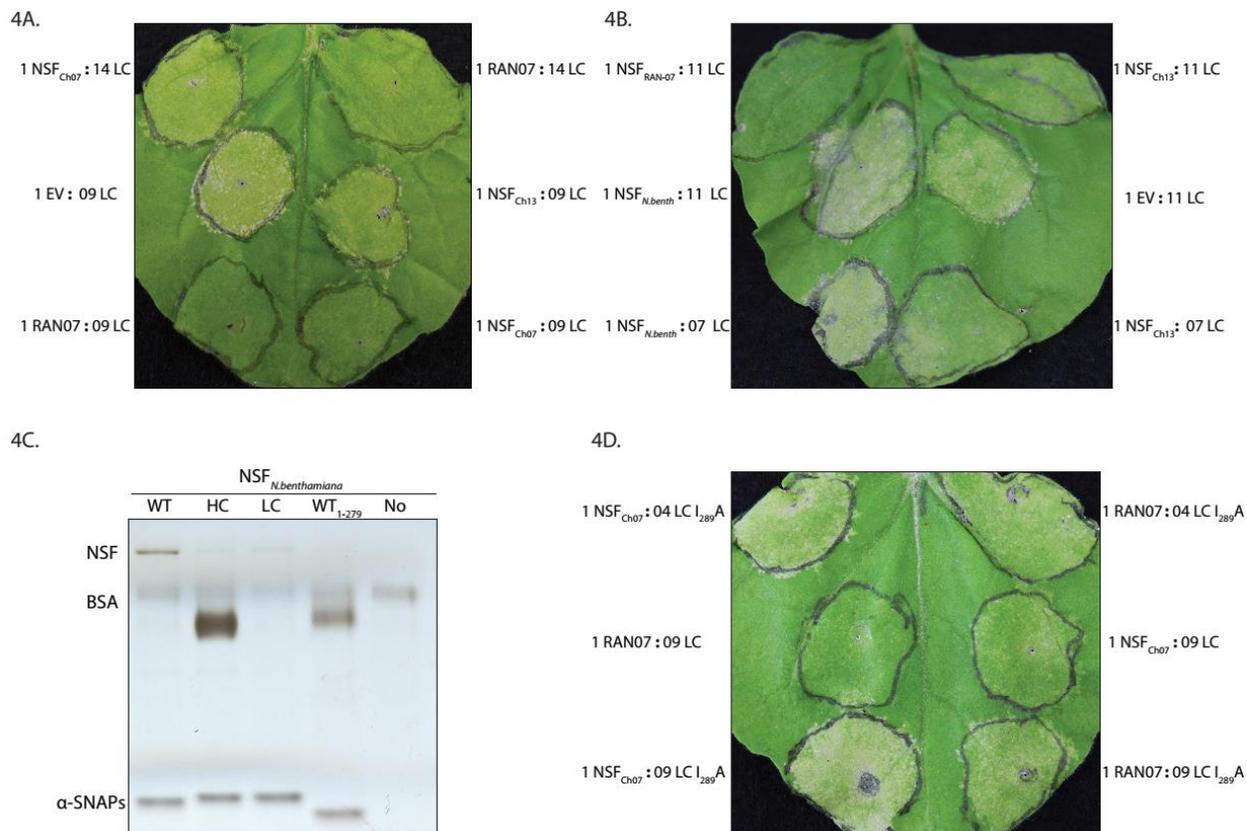
(A). Alignment of soybean NSF<sub>Ch07</sub>, NSF<sub>Ch13</sub>, and NSF<sub>RAN07</sub> N-terminal domains. Large identical regions omitted. N-domain residues which bind  $\alpha$ -SNAP colored red (N<sub>21</sub>, RR<sub>82-83</sub>, KK<sub>117-118</sub>).

NSF<sub>RAN07</sub> polymorphisms R<sub>4</sub>Q, S<sub>25</sub>N, <sub>116</sub>F, M<sub>181</sub>I colored green or purple (N<sub>21</sub>Y), unique NSF<sub>Ch13</sub> residues colored blue. (B) NSF<sub>RAN07</sub> modeled to NSF<sub>CHO</sub> cryo-EM structure (3J97A, State II). NSF residue patches implicated in  $\alpha$ -SNAP binding colored red and labeled I, II or III, respectively. (C) Like B, but NSF<sub>RAN07</sub> polymorphisms colored green or purple (N<sub>21</sub>Y), with zoomed in view of polymorphic N-domain region. (D). NSF N-domain R<sub>4</sub> is conserved in most model eukaryotes. Frequency logo of first 10 NSF N-domain residues of the following organisms: *Homo sapiens*, *Bos taurus*, *Mus musculus*, *Cricetulus griseus* (Chinese hamster), *Caenorhabditis elegans*, *Drosophila melanogaster*, *Danio rerio*, *Xenopus laevis*, *Gallus gallus*, *Neurospora crassa*, *Saccharomyces cerevisiae*, *Schizosaccharyomyces pombe*, *Chlamydomonas reinhardtii*, *Physcomitrella patens*, *Zea mays*, *Oryza sativa*, *Solanum tuberosum*, *Cucumis sativa*, *Arabidopsis thaliana*, *Medicago truncatula*, *Nicotiana benthamiana*, and *Glycine max*.



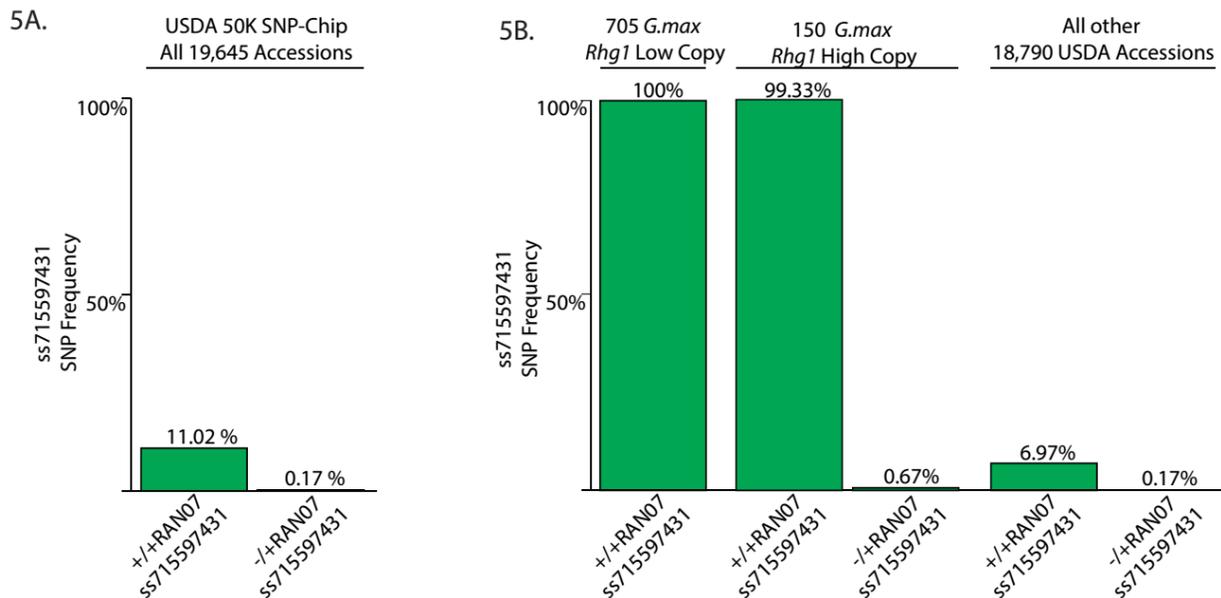
**Fig. 3.** NSF<sub>RAN07</sub> polymorphisms are at  $\alpha$ -SNAP binding interface and enhance binding with polymorphic *Rhg1*-resistance-type  $\alpha$ -SNAPs

(A). Cryo-EM structure of mammalian 20S supercomplex, masked to show only SNARE bundle (white), one  $\alpha$ -SNAP (yellow) and two NSF N-domains (light blue). Conserved NSF N-domain patches (I, R<sub>10</sub>; II, RK<sub>67-68</sub>; III, KK<sub>104-105</sub>) shown in red and  $\alpha$ -SNAP C-terminal contacts (D<sub>217</sub>DEED<sub>290-293</sub>) shown in orange. (B). NSF<sub>RAN07</sub> polymorphisms colored green, except N<sub>21</sub>Y in purple. (C). Silver-stained SDS/PAGE of recombinant NSF<sub>Ch07</sub> or NSF<sub>RAN07</sub> bound *in vitro* by the recombinant proteins indicated on second line: no- $\alpha$ -SNAP control (No) or wild-type (WT), low-copy (LC), or high copy (HC) *Rhg1*  $\alpha$ -SNAP. BSA: bovine serum albumin. (D) Densitometric quantification of NSF<sub>Ch07</sub> or NSF<sub>RAN07</sub> bound by *Rhg1*  $\alpha$ -SNAPs as in C; data are from three independent experiments and error bars show SEM.



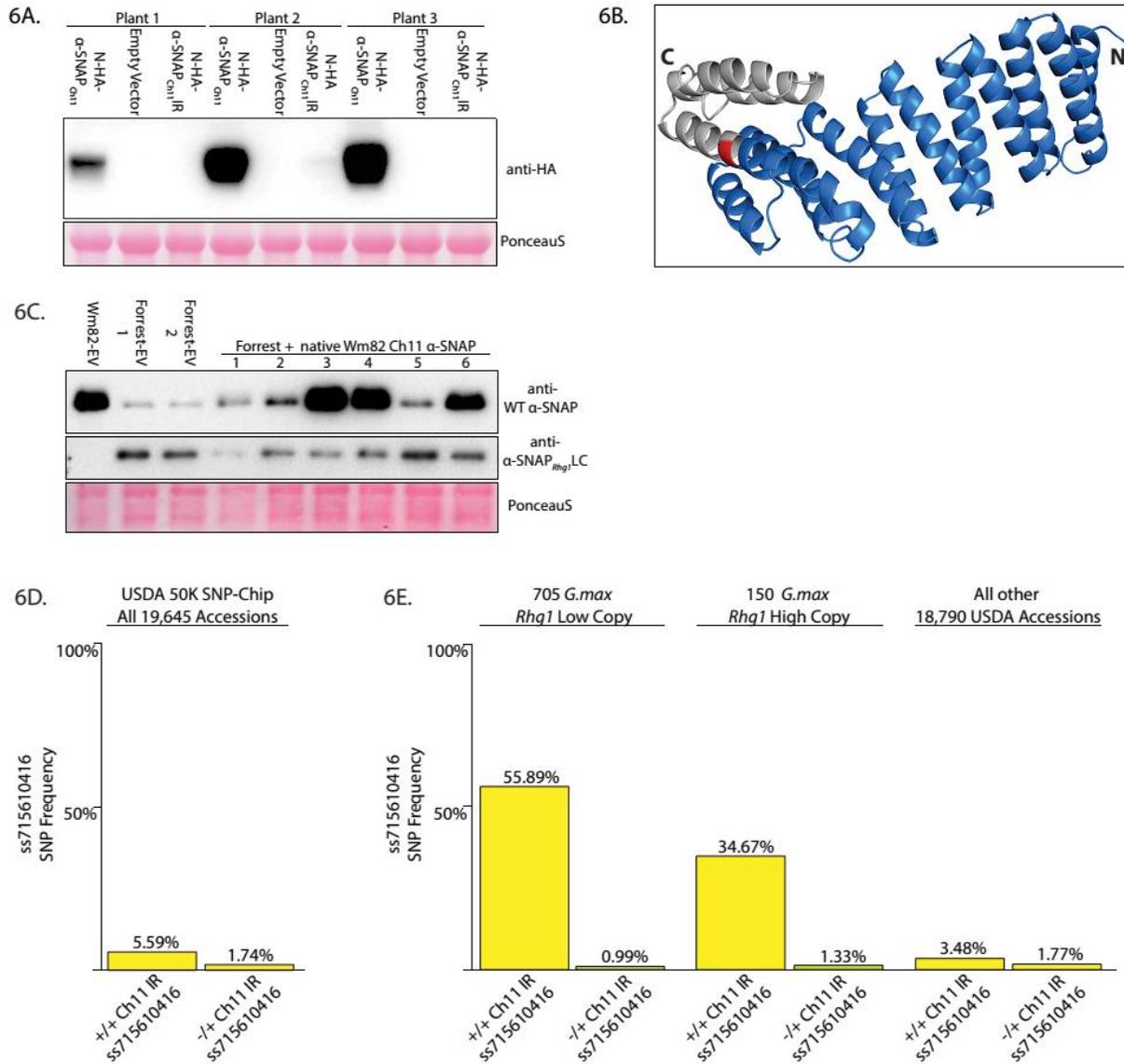
**Fig. 4.** Coexpression of soybean NSFs reduces cell-death symptoms caused by  $\alpha$ -SNAP<sub>Rhg1LC</sub>; NSF<sub>RAN07</sub> gives strongest protection.

(A). *N. benthamiana* leaves ~6 days post agro-infiltration with 9:1 or 14:1 mixed cultures of  $\alpha$ -SNAP<sub>Rhg1LC</sub> and NSF<sub>Ch07</sub> or NSF<sub>Ch13</sub> or NSF<sub>RAN07</sub> or empty vector (nine or fourteen parts *Agrobacterium* that delivers  $\alpha$ -SNAP<sub>Rhg1LC</sub> to one part *Agrobacterium* that delivers soybean NSF or empty vector control). (B). Like A, but 7:1 or 11:1 mixed cultures of  $\alpha$ -SNAP<sub>Rhg1LC</sub> co-expressed with NSF<sub>N.benth</sub> or NSF<sub>Ch13</sub> or NSF<sub>RAN07</sub> or empty vector. (C). Silver-stained SDS/PAGE of recombinant NSF<sub>N.benthamiana</sub> bound *in vitro* by recombinant wild-type, low-copy (LC), or high copy (HC) *Rhg1*  $\alpha$ -SNAP proteins or WT  $\alpha$ -SNAP lacking the final 10 C-terminal residues ( $\alpha$ -SNAP<sub>1-279</sub>). BSA, bovine serum albumin. (D). Like A and B, but 4:1 or 9:1 mixed cultures of  $\alpha$ -SNAP<sub>Rhg1LC</sub> or  $\alpha$ -SNAP<sub>Rhg1LC</sub>-I<sub>289A</sub> co-expressed with NSF<sub>Ch07</sub> or NSF<sub>RAN07</sub>.



**Fig. 5.** All *Rhg1*-signature soybeans in the USDA germplasms collection contain the R<sub>4</sub>Q NSF<sub>RAN07</sub> polymorphism

(A). Frequency of SoySNP50K SNP ss715597431 (corresponding to NSF<sub>RAN07</sub> R<sub>4</sub>Q) in all 19,645 SoySNP50K-genotyped *Glycine max* accessions. (B). Frequency of ss715597431 in all USDA *G. max* with *Rhg1*<sub>LC</sub> or *Rhg1*<sub>HC</sub> haplotype signatures or in remainder of SoySNP50K-genotyped *G. max* from USDA collection.



**Fig. 6.** The encoded  $\alpha$ -SNAP<sub>Ch11</sub>Intron Retention protein, unlike the WT  $\alpha$ -SNAP<sub>Ch11</sub>, is unstable.

(A) anti-HA immunoblot of *N. benthamiana* leaves agroinfiltrated to express empty vector, N-HA- $\alpha$ -SNAP<sub>Ch11</sub> or N-HA- $\alpha$ -SNAP<sub>Ch11</sub>-IR (intron-retention). PonceauS staining indicates relative total protein levels. (B) Modeling of  $\alpha$ -SNAP<sub>Ch11</sub> to sec17 crystal structure (yeast  $\alpha$ -SNAP, PDB ID 1QQE) suggests early termination of alpha-helix 12 in the intron-retention mutant. Termination point shown red, truncated residues shown grey. (C) Immunoblots for total

WT  $\alpha$ -SNAP and  $\alpha$ -SNAP<sub>*Rhg1*LC</sub> levels in Forrest (*Rhg1*LC) transgenic roots transformed with an empty vector (EV) or the native WT  $\alpha$ -SNAP<sub>Ch11</sub> locus from Williams 82. (D). Like 5A, except frequency of SoySNP50K SNP ss715610416 allele that is closest marker for  $\alpha$ -SNAP<sub>Ch11</sub>-IR, in all 19,645 USDA accessions. (E). Frequency of ss715610416 in all USDA *Glycine max* with *Rhg1*LC or *Rhg1*HC haplotype signatures vs. remainder of SoySNP50K-genotyped USDA collection.

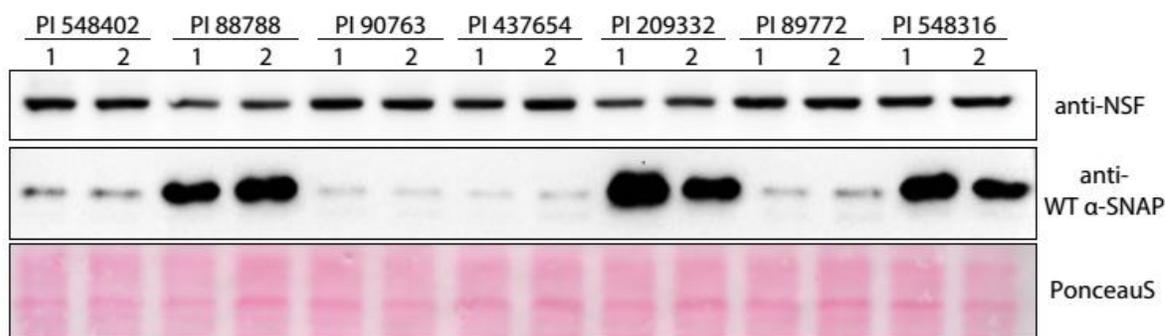
## 5.8.Tables

Line	<i>Rhg1</i> Haplotype	NSF <sub>Ch07</sub>	NSF <sub>Ch13</sub>
<b>Peking</b>	<i>Rhg1</i> <sub>LC</sub>	<i>Rhg1</i> Assoc. Allele	WT (Wm82-type)
<b>90763</b>	<i>Rhg1</i> <sub>LC</sub>	<i>Rhg1</i> Assoc. Allele	V555I
<b>437654</b>	<i>Rhg1</i> <sub>LC</sub>	<i>Rhg1</i> Assoc. Allele	WT (Wm82-type)
<b>209332</b>	<i>Rhg1</i> <sub>HC</sub>	<i>Rhg1</i> Assoc. Allele	V555I
<b>89772</b>	<i>Rhg1</i> <sub>LC</sub>	<i>Rhg1</i> Assoc. Allele	V555I
<b>548316</b>	<i>Rhg1</i> <sub>HC</sub>	<i>Rhg1</i> Assoc. Allele	V555I
<b>Prohio</b>	Susceptible	WT (Wm82-type)	V555I
<b>NE3001</b>	Susceptible	WT (Wm82-type)	Y260F
<b>4J105-34</b>	<i>Rhg1</i> <sub>HC</sub>	<i>Rhg1</i> Assoc. Allele	V555I, L738F
<b>CL0J095-46</b>	<i>Rhg1</i> <sub>HC</sub>	<i>Rhg1</i> Assoc. Allele	V555I
<b>IA3023</b>	Susceptible	WT (Wm82-type)	V555I
<b>LD00-3309</b>	<i>Rhg1</i> <sub>HC</sub>	<i>Rhg1</i> Assoc. Allele	WT (Wm82-type)
<b>LD02-4485</b>	<i>Rhg1</i> <sub>HC</sub>	<i>Rhg1</i> Assoc. Allele	WT (Wm82-type)
<b>LG05-4292</b>	<i>Rhg1</i> <sub>HC</sub>	<i>Rhg1</i> Assoc. Allele	WT (Wm82-type)
<b>LD01-5907</b>	<i>Rhg1</i> <sub>LC</sub>	<i>Rhg1</i> Assoc. Allele	V555I
<b>LD02-9050</b>	<i>Rhg1</i> <sub>HC</sub>	<i>Rhg1</i> Assoc. Allele	V555I
<b>Magellan</b>	Susceptible	WT (Wm82-type)	WT (Wm82-type)
<b>Maverick</b>	<i>Rhg1</i> <sub>HC</sub>	<i>Rhg1</i> Assoc. Allele	V555I

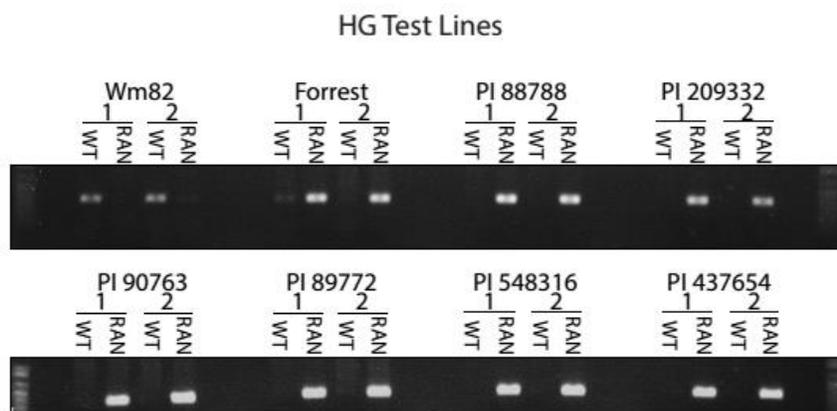
**Table 1:** HG type test lines and *Rhg1*-containing NAM Parents with a multi-copy *Rhg1* haplotype contain a unique NSF<sub>Ch07</sub> allele - *Rhg1* associated NSF on chromosome 07 (RAN07).

## 5.9 Supporting Information

S1A.



S1C.



**Fig. S1.** Wild-type  $\alpha$ -SNAP expression is reduced in *Rhgl*<sub>LowCopy</sub> soybeans; RAN07 is present within all examined *Rhgl* HG Type lines

(A). Independent immunoblot like Fig. 1A and incorporated into NSF densitometric analyses shown in Fig. 1B. Immunoblot of wild-type  $\alpha$ -SNAPs and NSF expression in HG type test soybean roots. *Rhgl*<sub>LC</sub> varieties: PI 548402 (Peking), PI 89772, PI 437654, PI 90763; *Rhgl*<sub>HC</sub> varieties: PI 88788, PI 209332, PI 548316(7 copy). PonceauS staining shows total protein loaded per lane. (B) Agarose gel showing PCR amplicons generated with RAN07 or NSF Ch07 WT specific primers on HG type soybeans and soybean genome reference variety Williams82 (Wm82). *Rhgl*<sub>LC</sub> varieties: “Forrest” (PI 548402-derived), PI 89772, PI 437654, PI 90763; *Rhgl*<sub>HC</sub> varieties: PI 88788, PI 209332, PI 548316 (7 copy).

NSF RAN07 alignment to Wild-Type NSF<sub>ch07</sub> (Wm82)

Wms82 MASRFGLSSSSSSASSMRVTNTPASDLALTNLAFCSPSDLRNFAVPGHNNLYLAAVADSF  
 RAN07 MASQFGLSSSSSSASSMRVYTPANDLALTNLAFCSPSDLRNFAVPGHNNLYLAAVADSF  
 \*\*\*:\*\*\*\*\* \*\*\*.\*\*\*\*\*

Wms82 VLSLSAHDTIGSGQIALNAVQRRCAKVSSGDSVQVSRFVPPEDFNLALLTLELEF VKKGS  
 RAN07 VLSLSAHDTIGSGQIALNAVQRRCAKVSSGDSVQVSRFVPPEDFNLALLTLELEF FVKKGS  
 \*\*\*\*\*

Wms82 KSEQIDAVLLAKQLRKRFRMNQVMTVGQKVLFEYHGNNYSFTVSNAAVEGQEKSNSLERGM  
 RAN07 KSEQIDAVLLAKQLRKRFRMNQVMTVGQKVLFEYHGNNYSFTVSNAAVEGQEKSNSLERGI  
 \*\*\*\*\*:

Wms82 ISDDTYIVFETSRDSGIKIVNQREGATSNI FKQKEFNLQSLGIGGLSAEFADI FRRAFAS  
 RAN07 ISDDTYIVFETSRDSGIKIVNQREGATSNI FKQKEFNLQSLGIGGLSAEFADI FRRAFAS  
 \*\*\*\*\*

Wms82 RVFPPHVTSKLGIKHVKGMLLYGPPGTGKTLMARQIGKILNGKEPKIVNGPEVLSKFVGE  
 RAN07 RVFPPHVTSKLGIKHVKGMLLYGPPGTGKTLMARQIGKILNGKEPKIVNGPEVLSKFVGE  
 \*\*\*\*\*

Wms82 TEKNVRDLFADAEQDQTRGDESDLHVI IFDEIDAICKSRGSTRDGTGVHDS IVNQLLTK  
 RAN07 TEKNVRDLFADAEQDQTRGDESDLHVI IFDEIDAICKSRGSTRDGTGVHDS IVNQLLTK  
 \*\*\*\*\*

Wms82 IDGVESLNNVLLIGMTNRKMDLDEALLRPGRLEVQVEISLPDENGLRQLIQLIHTNKMKEN  
 RAN07 IDGVESLNNVLLIGMTNRKMDLDEALLRPGRLEVQVEISLPDENGLRQLIQLIHTNKMKEN  
 \*\*\*\*\*

Wms82 SFLAADVNLQELAARTKNYSGAELEGVVKSAVS YALNRQLSLEDLTKPVEEENIKVTMDD  
 RAN07 SFLAADVNLQELAARTKNYSGAELEGVVKSAVS YALNRQLSLEDLTKPVEEENIKVTMDD  
 \*\*\*\*\*

Wms82 FLNALHEVTSAFGASTDDLERCRLHGMVECGDRHKHI YQRAMLLVEQVKVSKGSPLVTCCL  
 RAN07 FLNALHEVTSAFGASTDDLERCRLHGMVECGDRHKHI YQRAMLLVEQVKVSKGSPLVTCCL  
 \*\*\*\*\*

Wms82 LEGSRGSGKTALSATVGI DSDFPYVKIVSAESMIGLHESTKCAQIIKVFEDAYKSPLSVI  
 RAN07 LEGSRGSGKTALSATVGI DSDFPYVKIVSAESMIGLHESTKCAQIIKVFEDAYKSPLSVI  
 \*\*\*\*\*

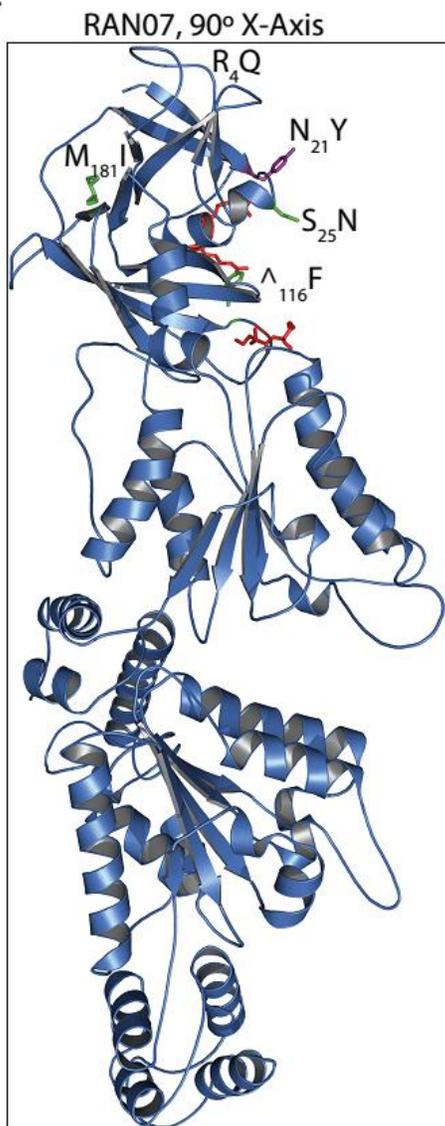
Wms82 ILDDIERLLEYVPIGPRFSNLISQTLVLLKRLPPKGKLMVIGTTSELDFLESIGFCDT  
 RAN07 ILDDIERLLEYVPIGPRFSNLISQTLVLLKRLPPKGKLMVIGTTSELDFLESIGFCDT  
 \*\*\*\*\*

Wms82 FSVTYHIPTLNTTDAKKVLEQLNVFTDEDIDSAAEALNDMP I RKL YMLIEMAAQGEHGGG  
 RAN07 FSVTYHIPTLNTTDAKKVLEQLNVFTDEDIDSAAEALNDMP I RKL YMLIEMAAQGEHGGG  
 \*\*\*\*\*

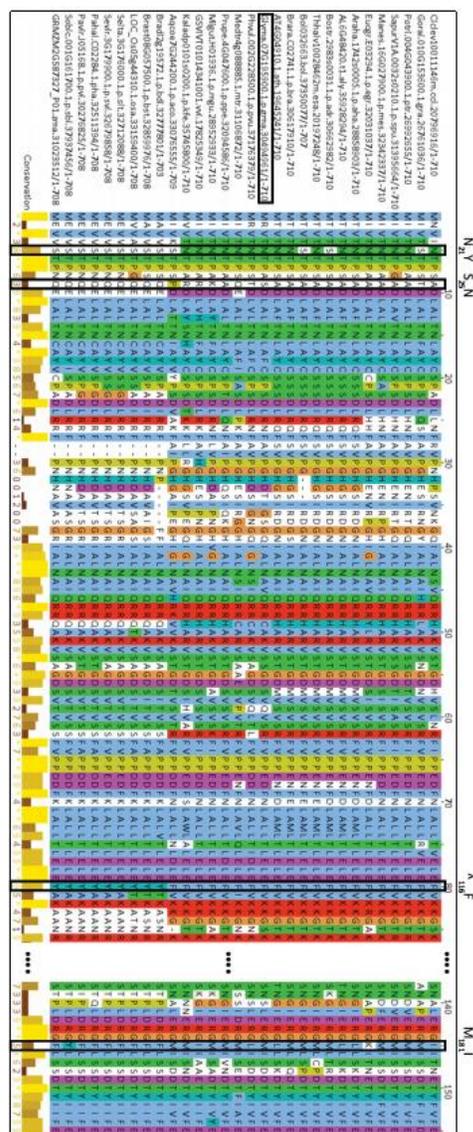
Wms82 AEAFSGKEKISIAHFYDCLQDVVRL  
 RAN07 AEAFSGKEKISIAHFYDCLQDVVRL  
 \*\*\*\*\*

**Fig. S2.** NSF RAN07 amino acid alignment with NSF<sub>Ch07</sub> of soybean reference genome Williams82. N-domain amino acid polymorphisms unique to RAN07 shown red. Corresponding residues in Wm82 NSF<sub>Ch07</sub> shown boldface.

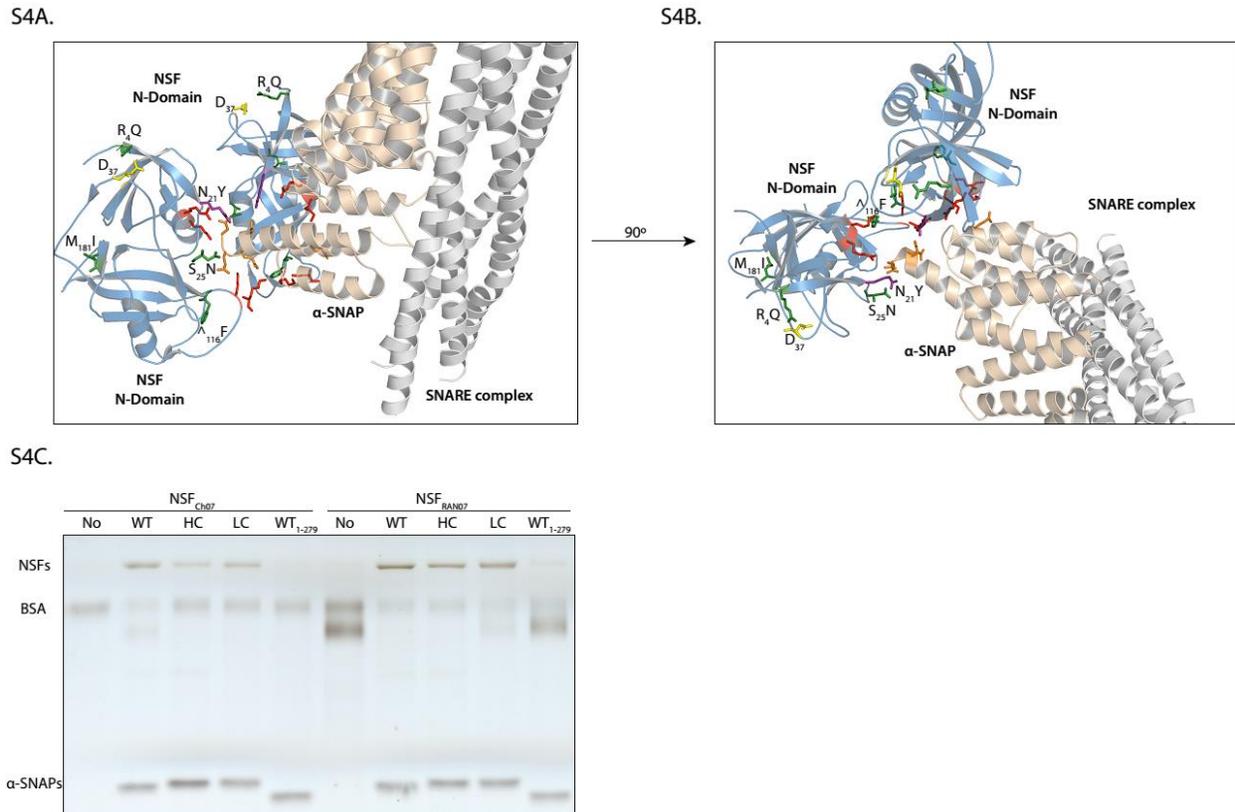
S3A.



S3B.

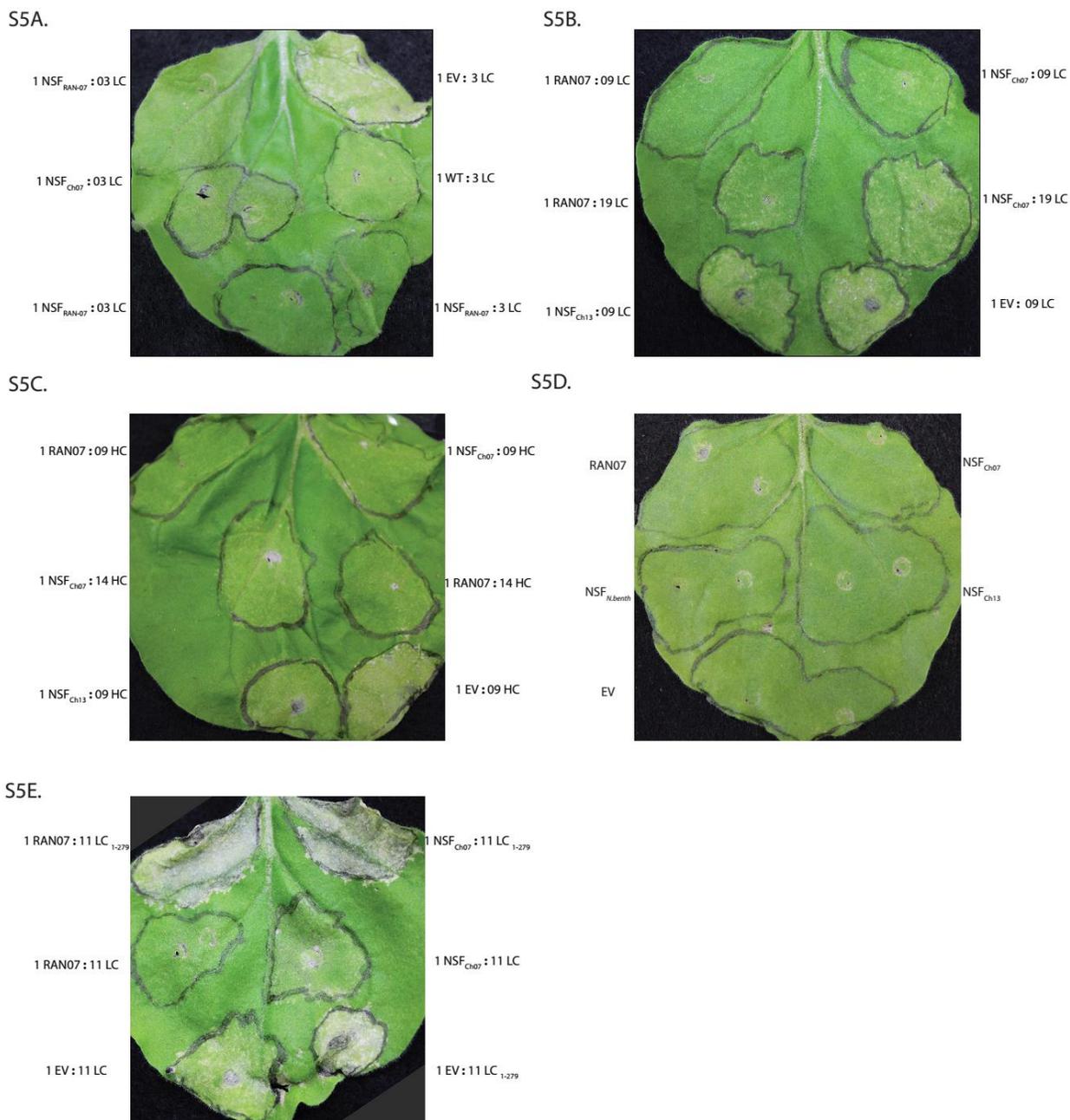
**Fig. S3.**

(A) NSF<sub>RAN07</sub> modeled to NSF<sub>CHO</sub> cryo-EM structure as in Fig. 2A, but rotated 90° on X-axis. NSF residue patches implicated in  $\alpha$ -SNAP binding colored red and labeled I, II or III, respectively. (B). Alignment of NSF N-domain using available plant NSF amino acid sequences from Phytozome.org. Alignment generated with Jalview starting at a conserved methionine residue corresponding to RAN07 met 17. Residues polymorphic in RAN07 are outlined with a box with the corresponding position labeled above.



**Fig. S4.** NSF<sub>RAN07</sub> polymorphisms are at  $\alpha$ -SNAP binding interface - R<sub>4</sub>Q positions near a conserved aspartate. RAN07 and NSF<sub>Ch07</sub> binding with  $\alpha$ -SNAP is dependent on the final 10  $\alpha$ -SNAP C-terminal residues.

(A). Like Fig. 4A, cryo-EM structure of mammalian 20S supercomplex showing SNARE bundle (white), one  $\alpha$ -SNAP (yellow) and two NSF N-domains (light blue). Conserved NSF N-domain patches (R<sub>10</sub>; RK<sub>67-68</sub>; KK<sub>104-105</sub>) shown red,  $\alpha$ -SNAP C-terminal contacts (D<sub>217</sub>DEED<sub>290-293</sub>) shown orange, NSF<sub>RAN07</sub> polymorphisms colored green, except N<sub>21</sub>Y in purple, D<sub>28</sub> shown yellow. (B). Same as A, but rotated 90° on Y-axis. (C). Same as Fig. 3C, except recombinant NSF<sub>Ch07</sub> or NSF<sub>RAN07</sub> bound *in vitro* by no- $\alpha$ -SNAP control (No) or wild-type (WT), low-copy (LC), or high copy (HC) *Rhg1*  $\alpha$ -SNAP, or WT  $\alpha$ -SNAP truncated at final 10 residues (WT<sub>1-279</sub>). BSA: bovine serum albumin.



**Fig. S5.** Coexpression of soybean NSFs reduces cell-death symptoms caused by *Rhg1* resistance  $\alpha$ -SNAPs; NSF<sub>RAN07</sub> gives strongest protection, but requires the polymorphic  $\alpha$ -SNAP<sub>*Rhg1*</sub>LC C-terminus for cell death complementation

(A). Like Fig. 4A, *N. benthamiana* leaves ~6 days post agro-infiltration but with 1:3 or 3:1 mixed cultures of  $\alpha$ -SNAP<sub>*Rhg1*</sub>LC and NSF<sub>Ch07</sub> or NSF<sub>RAN07</sub> or  $\alpha$ -SNAP<sub>*Rhg1*</sub>WT or empty vector

(one or three parts *Agrobacterium* that delivers  $\alpha$ -SNAP<sub>Rhg1</sub>LC to one part *Agrobacterium* that delivers soybean NSF, or  $\alpha$ -SNAP<sub>Rhg1</sub>WT or empty vector control). (B). Like Fig. 4A, but with a 9:1 or 19:1 mixed culture of  $\alpha$ -SNAP<sub>Rhg1</sub>LC co-expressed with NSF<sub>Ch07</sub> or NSF<sub>RAN07</sub> or empty vector. (C). Also like Fig. 4A, but using  $\alpha$ -SNAP<sub>Rhg1</sub>HC instead of  $\alpha$ -SNAP<sub>Rhg1</sub>LC in the corresponding mixture cultures of NSF<sub>Ch07</sub> or NSF<sub>RAN07</sub> or empty vector. (D) *N. benthamiana* leaves ~6 days post agro-infiltration with 1:9 mixed cultures of NSF<sub>Ch07</sub> or NSF<sub>RAN07</sub> or NSF<sub>Ch13</sub> or NSF<sub>N.benth</sub> to empty vector (9 parts empty vector cultures to 1 part NSF expressing *Agrobacterium* culture). (E). Like Fig. 4A, but with a 11:1 mixed culture of  $\alpha$ -SNAP<sub>Rhg1</sub>LC or  $\alpha$ -SNAP<sub>Rhg1</sub>LC<sub>1-280</sub> (lacks the final 10 C-terminal residues) co-expressed with NSF<sub>Ch07</sub> or NSF<sub>RAN07</sub> or empty vector.

**Alignment of NSF N.benthamiana and WT NSF<sub>Ch07</sub>**

Ch07 MASRFGLSSSSSSASSMRVT**NT**TPASDLALTNLAFCSPSDLRNFAVPGHNNLYLAAVADSF  
 Nben MAGRFG-----SGASTMIVT**NT**PAKDLAYTNCAYCSPADLRNFLVPGSK-LAYGLIADAF  
 . .\*.\*\*\*:\* \*\*\*\*\*.\*\*\* \*\* \*:\*:\*\*:\*\*\*\*\* \*\*\* : \* . :\*\*:\*

Ch07 VLSLSAHDTIGSGQIALNAVQ**RR**CAKVVSSGDSVQVSRFVPPEDFNLLLTLELEFV**KK**GS  
 Nben VLTAAHDGIPNGHLGLNAIQ**RR**YAKVSTGDTISVNRVFPDDFNLLLTIDLEFV**KK**GT  
 \*\*.\*:\*\* \* .\*::.\*\*\*:\*\* \* \*\*:\*:\*\*:..\*.\*\*\*\*\*:\*\*\*\*\*:..\*\*\*\*\*:

Ch07 KSEQIDAVLLAKQLRKRFMNQVMTVGQKVLFEYHGNNYSFTVSNAAVEGQEKSNSLERGM  
 Nben RDEQVDAVSLANQVRKKFANQIMSTGQKVTFEYHGNSYIFTVNQATVEGQEKSNI-IERGM  
 :.\*:\*\* \* \*\*:\*:\*\*:\* \*\*:\*:.\*\*\* \*\*\*\*\*.\* \*\*.\*:\*\*:\*\*\*\*\* :\*\*\*\*

Ch07 ISDDTYIVFETSRDSGIKIVNQREGATSNIFKQKEFNLSLQIGIGGLSAEFADIFRRAFAS  
 Nben ISADTYIIFEANSSGIKIVNQREAASSSIFRQKEFNLSLQIGIGGLSAEFADIFRRAFAS  
 \*\* \*\*\*\*\*:\*\*:..\*\*\*\*\*.\*:\*.\*:\*\*:\*\*\*\*\*:\*\*\*\*\*:\*\*\*\*\*

Ch07 RVFPPHVTSKLGIKHKVGMMLYGPPGTGKTLMARQIGKILNGKEPKIVNGPEVLSKRVGE  
 Nben RVFPPHVTSKLGIKHKVGMMLYGPPGTGKTLMARQIGKMLNGKEPKIVNGPEVLSKRVGE  
 \*\*\*\*\*:\*\*\*\*\*:\*\*\*\*\*

Ch07 TEKNVRDLFADAEQDQRTRGDES DLHVIIFDEIDAICKSRGSTRDGTGVHDSIVNQLLTK  
 Nben TEKNVRDLFADAEQDQRTKGDQSELHVIIFDEIDAICKSRGSTRDGTGVHDSIVNQLLTK  
 \*\*\*\*\*:\*\*\*:\*:\*\*:\*\*\*\*\*:\*\*\*\*\*:\*\*\*\*\*

Ch07 IDGVESLNNVLLIGMTNRKDMLDEALLRPGRLEVQVEISLPDENGRLQILQIHTNMKMN  
 Nben IDGVESLNNVLLIGMTNRKDLLDEALMRPGRLEVQVEISLPDENGRLQILQIHTNQMKEN  
 \*\*\*\*\*:\*\*\*\*\*:\*\*\*\*\*:\*\*\*\*\*:\*\*\*\*\*:\*\*\*\*\*

Ch07 SFLAADVNLQELAARTKNYSGAEELEGVVKSAVSALNRQLSLEDLTKPVEEENIKVTMDD  
 Nben SFLSPDVNLQELAARTKNYSGAEELEGVVKSAVSFALNRQLSMDDLTKPVEEESIKVTMDD  
 \*\*\*:..\*\*\*\*\*:\*\*\*\*\*:\*\*\*\*\*:\*\*\*.\*\*\*\*\*

Ch07 FLNALHEVTSAFGASTDDLERCRLHGMVECGDRHKHIYQRAMLLVEQVKVSKGSPLVTCL  
 Nben FLHALGEVRPAFGASTDDLERCRLNGIVDCGERHQHIYRRTMLLAEQVKVSRGSPLITCL  
 \*\*.\* \*\* .\*\*\*\*\*:\*\*\*:\*:\*\*:\*\*\*:\*\*\*:\*\*\*.\*\*\*\*\*:\*\*\*:\*\*\*

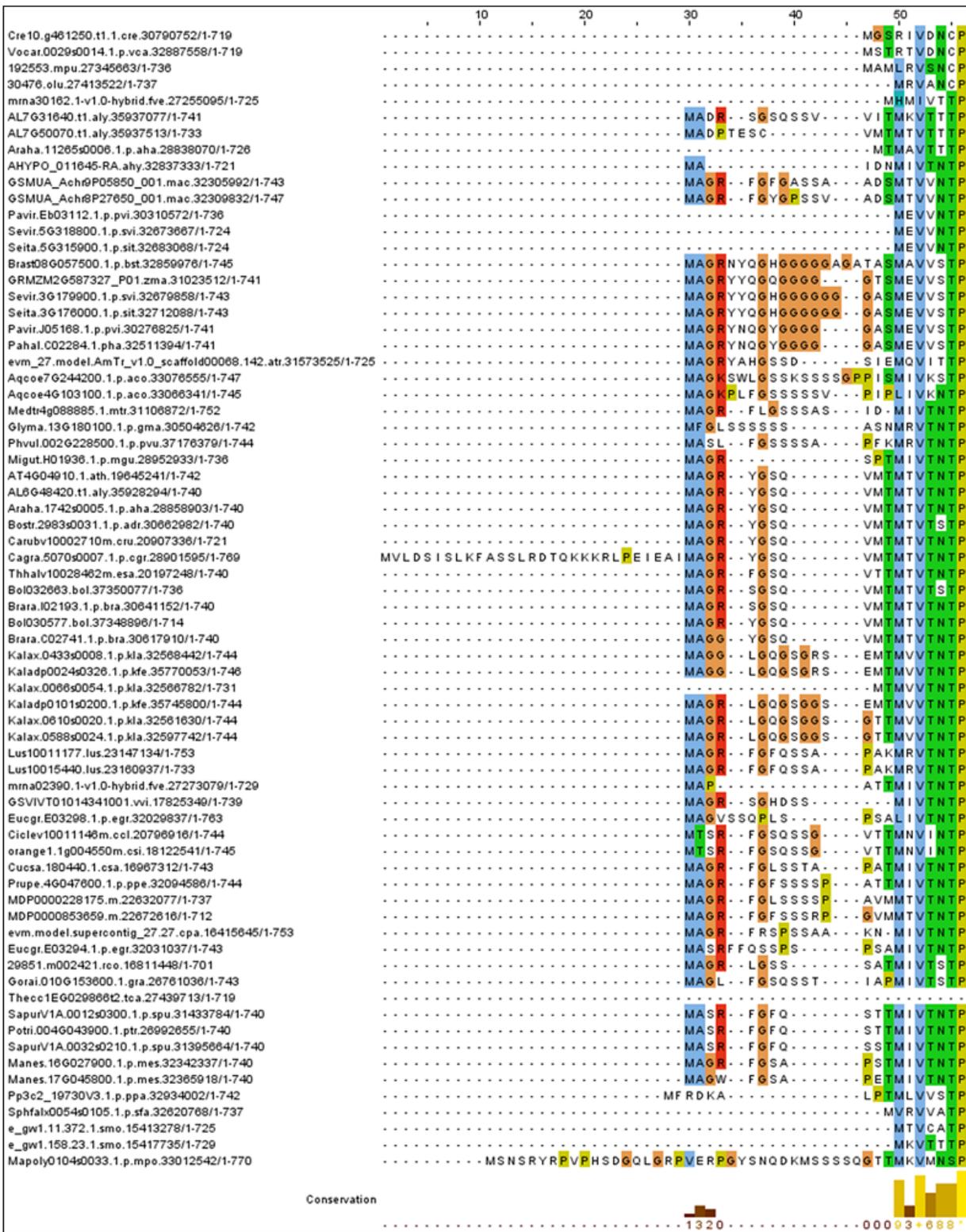
Ch07 LEGSRGSGKTALSATVGDSDFPYVKIVSAESMIGLHESTKCAQIIKVFEDAYKSPLSVI  
 Nben LEGPSGSGKTAMAATVGIESDFPYVKIISAETMIGLSESSKCAQIVKFEDAYKSPLSIV  
 \*\*\*. \*\*\*\*\*:\*\*\*\*\*:\*\*\*\*\*:\*\*\*:\*\*\* \*\*:\*:\*\*:\*\*\*\*\*:\*\*\*\*\*:..

Ch07 ILDDIERLLEYVPIGPRFSNLISQTLVLLKRLPPKGKMLVIGTTSELDFLESIGFCDT  
 Nben VLDGIERLLEYVAIGPRFSNLISQTLVLLKRLPPKGKILVIGTTSEAGFLDSVGLCDA  
 :\*.\*\*\*\*\*.\*\*\*\*\*:\*\*\*\*\*:\*\*\*\*\*.\*\*\*:\*:\*\*:\*\*\*

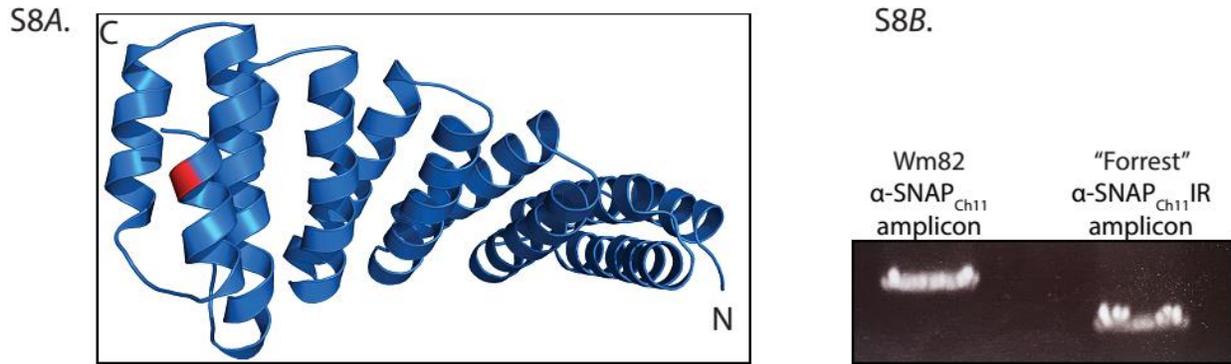
Ch07 FSVTYHIPTLNTTDAKKVLEQLNVFTDEDIDSAAEALNDMPKIRKLYMLIEMAAQGEHGG  
 Nben FSVTYHVPTLKTEDAKKVLQQLNVFSNDDVDSAAEALNDMPIKLYMVVEMAAQGEHGGT  
 \*\*\*\*\*:\*\*\*:\* \*\*\*\*\*:\*\*\*\*\*:..\*:\*\*:\*\*\*\*\*:\*\*\*\*\*:\*\*\*\*\*

Ch07 AEAI FSGKEKISIAHFYDCLQDVVRL  
 Nben AEAI YSGKEKIQISHFYDCLQDIARY  
 \*\*\*:\*\*\*\*\*.\*:\*\*:\*\*\*\*\*:..\*

**Fig. S6.** NSF *N. benthamiana* amino acid alignment with NSF<sub>Ch07</sub> of soybean reference genome Williams82. NSF N-domain residues conserved in  $\alpha$ -SNAP binding are shown red in boldface.



**Fig. S7.** Alignment of NSF N-domain starting from position 1 and showing general consensus of R<sub>4</sub>. Alignment generated with Jalview and includes all reliable (i.e., mostly complete sequences) Angiosperm NSF sequences available from Phytozome.org.

**Fig. S8.**

The encoded  $\alpha$ -SNAP<sub>Ch11</sub>Intron Retention protein may be unstable and undergoing pseudogenization. (A). Modeling of  $\alpha$ -SNAP<sub>Ch11</sub>-IR to sec17 crystal structure (yeast  $\alpha$ -SNAP, PDB ID 1QQE) suggests early termination of alpha-helix 12 (termination point shown red). (B). PCR across the putative promoter region of  $\alpha$ -SNAP<sub>Ch11</sub>-IR reveals an ~300 bp deletion not present in the Wm82  $\alpha$ -SNAP<sub>Ch11</sub> (WT) allele.

## 5.10 Supplemental Tables

Soybean Line	Glyma.07g195900	Glyma.07g195800	Glyma.07g195700	Glyma.07g195600	Glyma.07g195500	Glyma.07g195400	Glyma.07g195300	Glyma.07g195200	Glyma.07g195100
	NSF	Rubber Elongation Factor	DNA Mismatch Repair MutS2	No annotated domains	TFII H Polypeptide 4	E3 Ubiquitin Ligase	Asparagine Synthase	Uncharacterized Conserved Protein	LRR Containing Protein
PI 89772	R <sub>4</sub> Q, N <sub>21</sub> Y, S <sub>25</sub> N, <sup>^</sup> <sub>116</sub> F, M <sub>181</sub> I	K <sub>3</sub> N, F <sub>137</sub> S	T <sub>21</sub> A, K <sub>23</sub> R, G <sub>109</sub> C, H <sub>115</sub> Q, V <sub>345</sub> I, D <sub>364</sub> N, M <sub>406</sub> T, Q <sub>618</sub> K	WT	WT	WT	WT	WT	WT
PI 90763	R <sub>4</sub> Q, N <sub>21</sub> Y, S <sub>25</sub> N, <sup>^</sup> <sub>116</sub> F, M <sub>181</sub> I	K <sub>3</sub> N, L <sub>42</sub> R, F <sub>137</sub> S	T <sub>21</sub> A, K <sub>23</sub> R, G <sub>109</sub> C, H <sub>115</sub> Q, V <sub>345</sub> I, D <sub>364</sub> N, M <sub>406</sub> T, Q <sub>618</sub> K	WT	WT	WT	WT	WT	WT
PI 209332	R <sub>4</sub> Q, N <sub>21</sub> Y, S <sub>25</sub> N, <sup>^</sup> <sub>116</sub> F, M <sub>181</sub> I	K <sub>3</sub> N, L <sub>42</sub> R, F <sub>137</sub> S	T <sub>21</sub> A, K <sub>23</sub> R, V <sub>345</sub> I, D <sub>364</sub> N, M <sub>406</sub> T, Q <sub>618</sub> K	WT	WT	WT	WT	WT	WT
CLOJ095-4-6	R <sub>4</sub> Q, N <sub>21</sub> Y, S <sub>25</sub> N, <sup>^</sup> <sub>116</sub> F, M <sub>181</sub> I	K <sub>3</sub> N, L <sub>42</sub> R, F <sub>137</sub> S	T <sub>21</sub> A, K <sub>23</sub> R, G <sub>109</sub> C, H <sub>115</sub> Q, V <sub>345</sub> I, D <sub>364</sub> N, M <sub>406</sub> T, Q <sub>618</sub> K	WT	WT	WT	WT	WT	WT
IA3023	WT	L <sub>42</sub> R, F <sub>137</sub> S	D <sub>364</sub> N, M <sub>406</sub> T, Y <sub>576</sub> F	WT	WT	WT	E <sub>49</sub> G	D <sub>60</sub> A, S <sub>64</sub> P	WT
LD00-3309	R <sub>4</sub> Q, N <sub>21</sub> Y, S <sub>25</sub> N, <sup>^</sup> <sub>116</sub> F, M <sub>181</sub> I	K <sub>3</sub> N, L <sub>42</sub> R, F <sub>137</sub> S	T <sub>21</sub> A, K <sub>23</sub> R, G <sub>109</sub> C, H <sub>115</sub> Q, V <sub>345</sub> I, D <sub>364</sub> N, M <sub>406</sub> T, G <sub>516</sub> C, Q <sub>618</sub> K	WT	WT	WT	WT	WT	WT
PI 437654	R <sub>4</sub> Q, N <sub>21</sub> Y, S <sub>25</sub> N, <sup>^</sup> <sub>116</sub> F, M <sub>181</sub> I	K <sub>3</sub> N, F <sub>137</sub> S	T <sub>21</sub> A, K <sub>23</sub> R, G <sub>109</sub> C, H <sub>115</sub> Q, V <sub>345</sub> I, D <sub>364</sub> N, M <sub>406</sub> T, Q <sub>618</sub> K	WT	WT	WT	WT	WT	WT
PI 548402	R <sub>4</sub> Q, N <sub>21</sub> Y, S <sub>25</sub> N, <sup>^</sup> <sub>116</sub> F, M <sub>181</sub> I	K <sub>3</sub> N, F <sub>137</sub> S	T <sub>21</sub> A, K <sub>23</sub> R, G <sub>109</sub> C, H <sub>115</sub> Q, V <sub>345</sub> I, D <sub>364</sub> N, M <sub>406</sub> T, Q <sub>618</sub> K	WT	WT	WT	WT	WT	WT
Magellan	WT	L <sub>42</sub> R, F <sub>137</sub> S	D <sub>364</sub> N, M <sub>406</sub> T	WT	WT	WT	E <sub>49</sub> G	D <sub>60</sub> S, S <sub>64</sub> P	WT
Maverick	R <sub>4</sub> Q, N <sub>21</sub> Y, S <sub>25</sub> N, <sup>^</sup> <sub>116</sub> F, M <sub>181</sub> I	K <sub>3</sub> N, F <sub>137</sub> S	T <sub>21</sub> A, K <sub>23</sub> R, G <sub>109</sub> C, H <sub>115</sub> Q, V <sub>345</sub> I, D <sub>364</sub> N, M <sub>406</sub> T, Q <sub>618</sub> K	WT	WT	WT	WT	WT	WT
PI 548316	R <sub>4</sub> Q, N <sub>21</sub> Y, S <sub>25</sub> N, <sup>^</sup> <sub>116</sub> F, M <sub>181</sub> I	K <sub>3</sub> N, F <sub>137</sub> S	T <sub>21</sub> A, K <sub>23</sub> R, G <sub>109</sub> C, H <sub>115</sub> Q, V <sub>345</sub> I, D <sub>364</sub> N, M <sub>406</sub> T, Q <sub>618</sub> K	WT	WT	WT	WT	WT	WT

Table S1.

Amino acid polymorphisms of genes within the chromosome 07 interval co-segregating with *Rhg1*. Polymorphisms are compared to the predicted residue sequence of the Wm82 (SCN-susceptible) reference genome. The predicted amino acid sequence of most candidate loci matches Wm82 (SCN-susceptible) sequence. Among candidate loci with amino acid substitutions, only RAN07 (shown light blue) has the same consistent amino acid changes across all examined *Rhg1*-containing germplasm. SCN-susceptible soybean varieties colored green.

Diverse Parent	RR (Ch07, Ch18)	RS(Ch07, Ch18)	SR(Ch07, Ch18)	SS(Ch07, Ch18)	HR(Ch07, Ch18)	HS(Ch07, Ch18)	HH(Ch07, Ch18)	RH(Ch07, Ch18)	SH(Ch07, Ch18)
4J105-3-4	41	41	2	31	9	3	1	9	0
CLOJ095-4-6	35	45	0	37	6	7	0	7	1
LD00-3309	38	45	1	27	8	10	3	7	0
LD01-5907	32	32	1	42	0	6	1	6	2
LD02-4485	37	50	1	28	10	7	1	5	0
LD02-9050	43	31	2	34	10	10	1	4	0
Maverick	31	34	0	41	8	8	3	8	1
LG05-4292	44	41	1	30	1	3	0	7	0
Totals	301	319	8*	270	52	54	10	53	4

R refers to allele from *Rhg1* resistant parent.

S refers to allele from SCN-susceptible parent

Genotype order: first allele is chr 7 (RAN07 interval) and second is chr 18 (*Rhg1* interval)

\*All 8 re-examined RILs that inherited *Rhg1*<sub>HC</sub> or *Rhg1*<sub>LC</sub> also inherited the *NSF*<sub>RAN07</sub><sup>^116 F</sup> and *M*<sub>181</sub>I mutations meaning that all 309 RILs that carried the resistance associated *Rhg1* also carried *NSF*<sub>RAN07</sub>

**Table S2.** *NSF*<sub>RAN07</sub> co-segregates with *Rhg1* in all *Rhg1*-containing F<sub>2.5</sub> offspring from *Rhg1*<sup>+</sup> x *rhg1*<sup>-</sup> crosses.

## 5.11 References

- Allen TW, Bradley, C.A., Damicone, J.P., Dufault, N.S., Faske, T.R., Hollier, C.A., Isakeit T, Kemerait, R.C., Kleczewski, N.M., Kratochvil, R.J., Mehl, H.L., Mueller, J.D. O, C., Price, P.P., Sikora, E.J., Spurlock, T.N., Thiessen, L., Wiebold, W.J. aY, H** (2016) Southern United States Soybean Disease Loss Estimates for 2016.
- Bachem CW, Oomen RJJ, Kuyt S, Horvath BM, Claassens MM, Vreugdenhil D, Visser RG** (2000) Antisense suppression of a potato alpha-SNAP homologue leads to alterations in cellular development and assimilate distribution. *Plant Mol Biol* **43**: 473-482
- Baker RW, Hughson FM** (2016) Chaperoning SNARE assembly and disassembly. *Nat Rev Mol Cell Biol* **17**: 465-479
- Barnard RJ, Morgan A, Burgoyne RD** (1996) Domains of alpha-SNAP required for the stimulation of exocytosis and for N-ethylmaleimide-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
- Bassham DC, Raikhel NV** (1999) The pre-vacuolar t-SNARE AtPEP12p forms a 20S complex that dissociates in the presence of ATP. *Plant J* **19**: 599-603
- Bayless AM, Smith JM, Song J, McMinn PH, Teillet A, August BK, Bent AF** (2016) Disease resistance through impairment of alpha-SNAP-NSF interaction and vesicular trafficking by soybean Rhg1. *Proc Natl Acad Sci U S A* **113**: E7375-E7382
- Bekal S, Domier LL, Gonfa B, Lakhssassi N, Meksem K, Lambert KN** (2015) A SNARE-Like Protein and Biotin Are Implicated in Soybean Cyst Nematode Virulence. *PLoS One* **10**: e0145601
- 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. *Theor Appl Genet* **111**: 44-49
- Chae TH, Kim S, Marz KE, Hanson PI, Walsh CA** (2004) The hyh mutation uncovers roles for alpha Snap in apical protein localization and control of neural cell fate. *Nat Genet* **36**: 264-270
- Concibido VC, Diers BW, Arelli PR** (2004) A Decade of QTL Mapping for Cyst Nematode Resistance in Soybean. *Crop Sci.* **44**: 1121-1131

- Cook DE, Bayless AM, Wang K, Guo X, Song Q, Jiang J, Bent AF** (2014) Distinct Copy Number, Coding Sequence, and Locus Methylation Patterns Underlie Rhg1-Mediated Soybean Resistance to Soybean Cyst Nematode. *Plant Physiol* **165**: 630-647
- 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
- Crooks GE, Hon G, Chandonia JM, Brenner SE** (2004) WebLogo: a sequence logo generator. *Genome Res* **14**: 1188-1190
- 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. *Nat Genet* **43**: 491-498
- Dodds PN, Rathjen JP** (2010) Plant immunity: towards an integrated view of plant-pathogen interactions. *Nat Rev Genet* **11**: 539-548
- Gheysen G, Mitchum MG** (2011) How nematodes manipulate plant development pathways for infection. *Curr Opin Plant Biol* **14**: 415-421
- Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS** (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res* **40**: D1178-1186
- Grant D, Nelson RT, Cannon SB, Shoemaker RC** (2010) SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Res* **38**: D843-846
- Griff IC, Schekman R, Rothman JE, Kaiser CA** (1992) The yeast SEC17 gene product is functionally equivalent to mammalian alpha-SNAP protein. *J Biol Chem* **267**: 12106-12115
- Hanson PI, Whiteheart SW** (2005) AAA+ proteins: have engine, will work. *Nat Rev Mol Cell Biol* **6**: 519-529
- Hewezi T, Baum TJ** (2013) Manipulation of plant cells by cyst and root-knot nematode effectors. *Mol Plant Microbe Interact* **26**: 9-16
- Horsnell WG, Steel GJ, Morgan A** (2002) Analysis of NSF mutants reveals residues involved in SNAP binding and ATPase stimulation. *Biochemistry* **41**: 5230-5235
- Jahn R, Scheller RH** (2006) SNAREs--engines for membrane fusion. *Nat Rev Mol Cell Biol* **7**: 631-643

- Jones JT, Haegeman A, Danchin EG, Gaur HS, Helder J, Jones MG, Kikuchi T, Manzanilla-Lopez R, Palomares-Rius JE, Wesemael WM, Perry RN** (2013) Top 10 plant-parasitic nematodes in molecular plant pathology. *Mol Plant Pathol* **14**: 946-961
- Kopisch-Obuch FJ, Diers BW** (2006) Segregation at the SCN resistance locus *rhg1* in soybean is distorted by an association between the resistance allele and reduced field emergence. *Theor Appl Genet* **112**: 199-207
- Kyndt T, Vieira P, Gheysen G, de Almeida-Engler J** (2013) Nematode feeding sites: unique organs in plant roots. *Planta* **238**: 807-818
- Lakhssassi N, Liu S, Bekal S, Zhou Z, Colantonio V, Lambert K, Barakat A, Meksem K** (2017) Characterization of the Soluble NSF Attachment Protein gene family identifies two members involved in additive resistance to a plant pathogen. *Sci Rep* **7**: 45226
- Lee TG, Kumar I, Diers BW, Hudson ME** (2015) Evolution and selection of *Rhg1*, a copy-number variant nematode-resistance locus. *Mol Ecol* **24**: 1774-1791
- Li H, Durbin R** (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754-1760
- Littleton JT, Barnard RJ, Titus SA, Slind J, Chapman ER, Ganetzky B** (2001) SNARE-complex disassembly by NSF follows synaptic-vesicle fusion. *Proc Natl Acad Sci U S A* **98**: 12233-12238
- Liu S, Kandoth PK, Lakhssassi N, Kang J, Colantonio V, Heinz R, Yeckel G, Zhou Z, Bekal S, Dapprich J, Rotter B, Cianzio S, Mitchum MG, Meksem K** (2017) The soybean *GmSNAP18* gene underlies two types of resistance to soybean cyst nematode. *Nat Commun* **8**: 14822
- 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
- Lobingier BT, Nickerson DP, Lo SY, Merz AJ** (2014) SM proteins *Sly1* and *Vps33* co-assemble with *Sec17* and SNARE complexes to oppose SNARE disassembly by *Sec18*. *Elife* **3**: e02272
- 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
- Miao Y, Miner C, Zhang L, Hanson PI, Dani A, Vig M** (2013) An essential and NSF independent role for alpha-SNAP in store-operated calcium entry. *Elife* **2**: e00802

- Mitchum MG** (2016) Soybean Resistance to the Soybean Cyst Nematode *Heterodera glycines*: An Update. *Phytopathology* **106**: 1444-1450
- Mitchum MG, Hussey RS, Baum TJ, Wang X, Elling AA, Wubben M, Davis EL** (2013) Nematode effector proteins: an emerging paradigm of parasitism. *New Phytol*
- Mitchum MG, Sukno S, Wang X, Shani Z, Tsabary G, Shoseyov O, Davis EL** (2004) The promoter of the *Arabidopsis thaliana* Cell endo-1,4-beta glucanase gene is differentially expressed in plant feeding cells induced by root-knot and cyst nematodes. *Mol Plant Pathol* **5**: 175-181
- Naydenov NG, Harris G, Brown B, Schaefer KL, Das SK, Fisher PB, Ivanov AI** (2012) Loss of soluble N-ethylmaleimide-sensitive factor attachment protein alpha (alphaSNAP) induces epithelial cell apoptosis via down-regulation of Bcl-2 expression and disruption of the Golgi. *J Biol Chem* **287**: 5928-5941
- Niblack TL, Arelli PR, Noel GR, Opperman CH, Orf JH, Schmitt DP, Shannon JG, Tylka GL** (2002) A Revised Classification Scheme for Genetically Diverse Populations of *Heterodera glycines*. *J Nematol* **34**: 279-288
- Niblack TL, Lambert KN, Tylka GL** (2006) A model plant pathogen from the kingdom Animalia: *Heterodera glycines*, the soybean cyst nematode. *Annu Rev Phytopathol* **44**: 283-303
- Rancour DM, Dickey CE, Park S, Bednarek SY** (2002) Characterization of AtCDC48. Evidence for multiple membrane fusion mechanisms at the plane of cell division in plants. *Plant Physiol* **130**: 1241-1253
- 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
- Sanderfoot AA, Assaad FF, Raikhel NV** (2000) The *Arabidopsis* genome. An abundance of soluble N-ethylmaleimide-sensitive factor adaptor protein receptors. *Plant Physiol* **124**: 1558-1569
- Sanyal S, Krishnan KS** (2001) Lethal comatose mutation in *Drosophila* reveals possible role for NSF in neurogenesis. *Neuroreport* **12**: 1363-1366
- Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, 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. *Nature* **463**: 178-183

- Song H, Orr AS, Duan M, Merz AJ, Wickner WT** (2017) Sec17/Sec18 act twice, enhancing membrane fusion and then disassembling cis-SNARE complexes. *Elife* **6**
- Song J, Keppler BD, Wise RR, Bent AF** (2015) PARP2 Is the Predominant Poly(ADP-Ribose) Polymerase in Arabidopsis DNA Damage and Immune Responses. *PLoS Genet* **11**: e1005200
- Song Q, Hyten DL, Jia G, Quigley CV, Fickus EW, Nelson RL, Cregan PB** (2013) Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. *PLoS One* **8**: e54985
- Song Q, Hyten DL, Jia G, Quigley CV, Fickus EW, Nelson RL, Cregan PB** (2015) Fingerprinting Soybean Germplasm and Its Utility in Genomic Research. *G3 (Bethesda)* **5**: 1999-2006
- Song Q, Yan L, Quigley C, Jordan BD, Fickus E, Schroeder S, Song BH, Charles An YQ, Hyten D, Nelson R, Rainey K, Beavis WD, Specht J, Diers B, Cregan P** (2017) Genetic Characterization of the Soybean Nested Association Mapping Population. *Plant Genome* **10**
- Sudhof TC, Rothman JE** (2009) Membrane fusion: grappling with SNARE and SM proteins. *Science* **323**: 474-477
- T. W. Allen CAB, A. J. Sisson, E. Byamukama, M. I. Chilvers, C. M. Coker, A. A. Collins, J. P. Damicone, A. E. Dorrance, N. S. Dufault, P. D. Esker, T. R. Faske, L. J. Giesler, A. P. Grybauskas, D. E. Hershman, C. A. Hollier, T. Isakeit, D. J. Jardine, H. M. Kelly, R. C. Kemerait, N. M. Kleczewski, S. R. Koenning, J. E. Kurle, D. K. Malvick, S. G. Markell, H. L. Mehl, D. S. Mueller, J. D. Mueller, R. P. Mulrooney, B. D. Nelson, M. A. Newman, L. Osborne, C. Overstreet, G. B. Padgett, P. M. Phipps, P. P. Price, E. J. Sikora, D. L. Smith, T. N. Spurlock, C. A. Tande, A. U. Tenuta, K. A. Wise, and J. A. Wrather.** (2017) Soybean Yield Loss Estimates Due to Diseases in the United States and Ontario, Canada, from 2010 to 2014. *Plant Health Research*
- Vuong TD, Sonah H, Meinhardt CG, Deshmukh R, Kadam S, Nelson RL, Shannon JG, Nguyen HT** (2015) Genetic architecture of cyst nematode resistance revealed by genome-wide association study in soybean. *BMC Genomics* **16**: 593
- Webb DM, Baltazar BM, Rao-Arelli AP, Schupp J, Clayton K, Keim P, Beavis WD** (1995) Genetic mapping of soybean cyst nematode race-3 resistance loci in the soybean PI 437.654. *Theor Appl Genet* **91**: 574-581
- Whiteheart SW, Matveeva EA** (2004) Multiple binding proteins suggest diverse functions for the N-ethylmaleimide sensitive factor. *J Struct Biol* **146**: 32-43

- Whiteheart SW, Schraw T, Matveeva EA** (2001) N-ethylmaleimide sensitive factor (NSF) structure and function. *Int Rev Cytol* **207**: 71-112
- Wickner W** (2010) Membrane fusion: five lipids, four SNAREs, three chaperones, two nucleotides, and a Rab, all dancing in a ring on yeast vacuoles. *Annu Rev Cell Dev Biol* **26**: 115-136
- Wickner W, Schekman R** (2008) Membrane fusion. *Nat Struct Mol Biol* **15**: 658-664
- Yu N, Lee TG, Rosa DP, Hudson M, Diers BW** (2016) Impact of Rhg1 copy number, type, and interaction with Rhg4 on resistance to *Heterodera glycines* in soybean. *Theor Appl Genet* **129**: 2403-2412
- Zhao C, Matveeva EA, Ren Q, Whiteheart SW** (2010) Dissecting the N-ethylmaleimide-sensitive factor: required elements of the N and D1 domains. *J Biol Chem* **285**: 761-772
- Zhao C, Slevin JT, Whiteheart SW** (2007) Cellular functions of NSF: not just SNAPs and SNAREs. *FEBS Lett* **581**: 2140-2149
- Zhao M, Brunger AT** (2015) Recent Advances in Deciphering the Structure and Molecular Mechanism of the AAA+ ATPase N-Ethylmaleimide-Sensitive Factor (NSF). *J Mol Biol*
- Zhao M, Brunger AT** (2016) Recent Advances in Deciphering the Structure and Molecular Mechanism of the AAA+ ATPase N-Ethylmaleimide-Sensitive Factor (NSF). *J Mol Biol* **428**: 1912-1926
- Zhao M, Wu S, Zhou Q, Vivona S, Cipriano DJ, Cheng Y, Brunger AT** (2015) Mechanistic insights into the recycling machine of the SNARE complex. *Nature* **518**: 61-67
- Zick M, Orr A, Schwartz ML, Merz AJ, Wickner WT** (2015) Sec17 can trigger fusion of trans-SNARE paired membranes without Sec18. *Proc Natl Acad Sci U S A* **112**: E2290-2297

**Chapter 6: An intact retrotransposon is integrated within the  $\alpha$ -SNAP-encoding gene of low-copy haplotypes of the *Rhg1* locus**

This chapter is in preparation for publication:

Adam Bayless, Ryan Zapotocny, Kaela Amundson, Andrew Bent

Contributions: I conceived of this project with Andrew Bent. I performed the majority of the described experiments and wrote the manuscript with Andrew Bent. Ryan Zapotocny performed methylation analysis of the Copia junction.

## 6.1 Abstract

The Resistance to *Heterodera glycines* 1 (*Rhg1*) locus is used worldwide to control the most economically damaging soybean pathogen - soybean cyst nematode (SCN). *Rhg1* is a four gene block that is tandemly repeated, and *Rhg1* haplotypes are categorized as *Rhg1* high copy (*Rhg1<sub>HC</sub>*, 4 or more repeats) or *Rhg1* low copy (*Rhg1<sub>LC</sub>*, 3 or less repeats) based on repeat copy number. Both *Rhg1* haplotypes encode unique  $\alpha$ -SNAP (alpha-Soluble NSF Attachment Protein) proteins which are impaired in normal N-ethylmaleimide Sensitive Factor (NSF) interactions, but play key roles during the SCN-resistance response. Here, we report that *Rhg1<sub>LC</sub>* haplotypes contain a previously unrecognized gene, making the locus larger than previously reported - a 4.77 kb Copia retrotransposon is integrated within the  $\alpha$ -SNAP locus. This Copia element is integrated within  $\alpha$ -SNAP intron 1, oriented anti-sense to the  $\alpha$ -SNAP open reading frame, and is spliced from the  $\alpha$ -SNAP mRNA. All examined *Rhg1<sub>LC</sub>*  $\alpha$ -SNAP encoding repeats harbor this Copia integration, which we termed RAC (*Rhg1<sub>LC</sub>*  $\alpha$ -SNAP Copia). RAC was not detected in the  $\alpha$ -SNAPs encoded by *Rhg1<sub>HC</sub>* or SCN-susceptible soybeans. DNA methylation was detected at the  $\alpha$ -SNAP-RAC junctions, but small RNA reads aligning to RAC were not abundant. Additionally, mRNA transcripts containing RAC-sequence were detected from native *Rhg1<sub>LC</sub>*  $\alpha$ -SNAP locus. We interrogated if silencing RAC or removing RAC from the native  $\alpha$ -SNAP locus affected  $\alpha$ -SNAP protein abundance, but no impacts were observed. Most commercial soybeans utilize an *Rhg1<sub>HC</sub>* haplotype and overuse of this source is selecting for virulent SCN populations. Use of *Rhg1<sub>LC</sub>* as an alternate SCN resistance source is of interest to growers, and our findings indicate that *Rhg1<sub>LC</sub>* haplotypes contain a potentially activate retrotransposon, which could impact expression of the key *Rhg1*  $\alpha$ -SNAP locus. This study indicates that the two resistance-conferring multi-copy *Rhg1* haplotypes are more divergent than previously recognized.

## 6.2 Introduction

Transposable elements (TEs) constitute large percentages of plant and animal genomes, with the rice and maize genomes consisting of ~40% and 80% TEs, respectively (Meyers et al., 2001; Du et al., 2010; Mita and Boeke, 2016). In plants, transposons belonging to the Long Terminal Repeat (LTR) of retrotransposons are particularly abundant (Zhao and Ma, 2013; Galindo-Gonzalez et al., 2017). TEs and LTR retrotransposons have often been considered “junk DNA” or genomic parasites given their sequence similarities to retroviruses and often apparent absence of benefit (Havecker et al., 2004; Schulman, 2013; Zhao and Ma, 2013; Mita and Boeke, 2016). LTR retrotransposons are classified into two main superfamilies - Copia or Gypsy - based on gene order. However, both replicate using RNA intermediates which are reverse transcribed and then integrated within the host chromosome (Havecker et al., 2004; Schulman, 2013). If TE replication goes unchecked, TE numbers can quickly increase and the chances of a deleterious TE integration occurring within or adjacent to important host gene rises (Kidwell and Lisch, 2001; Lisch, 2009). Plants and animals therefore safeguard genomic integrity by silencing TEs, typically through RNA directed DNA methylation and small RNA silencing pathways (Lisch, 2009; Lisch and Slotkin, 2011).

To thrive in adverse environments, plants must respond accordingly with environmental changes, including abiotic and biotic stresses. Certain transposable element families reportedly activate during stress conditions and escape from normal host regulatory controls (Woodrow et al., 2010; Matsunaga et al., 2012; Cavrak et al., 2014; Matsunaga et al., 2015; Galindo-Gonzalez et al., 2017). Typically, *cis*-regulatory motifs (often within retrotransposon LTRs) within the TE can recruit certain stress-responsive host transcriptional factors (Galindo-Gonzalez et al., 2017) (to do what?). During these stressful periods, TEs can also directly influence nearby host gene

activation or effect alterations in the surrounding epigenetic landscape, potentially conferring host benefits during a stress (i.e., heat, salt, pathogen)(Slotkin and Martienssen, 2007; McCue and Slotkin, 2012; Galindo-Gonzalez et al., 2017). Such beneficial, regulatory transposons are referred to as “domesticated” transposons (Tsuchiya and Eulgem, 2013). Notably, domesticated transposons have been shown to have roles in plant defense gene expression (Tsuchiya and Eulgem, 2013). TEs have been shown to underlie many other plant host phenotypes, including but not limited to flowering response, trichomes and fruit size (Liu et al., 2004; Xiao et al., 2008; Ding et al., 2015)

*Glycine max* (soybean) is the world’s most widely farmed legume and a major source of food protein and oil (Schmutz et al., 2010). A major biotic stress affecting worldwide soybean production is an obligate soybean root parasite - the soybean cyst nematode (*Heterodera glycines*, SCN)(Niblack et al., 2006; Mitchum, 2016). SCN infestation can stunt growth and diminish seed production, thereby reducing both yield and fitness (Niblack et al., 2006; Mitchum, 2016). Soybeans maintain several natural defensive loci which restrict SCN success (Mitchum, 2016). Among these protective loci, the *Rhg1* (Resistance to *Heterodera glycines* 1) locus confers the greatest SCN resistance, and therefore is used in all commercially SCN resistant soybeans(Concibido et al., 2004).

*Rhg1* is an atypical plant disease resistance locus(Cook et al., 2012). *Rhg1* is a copy number variant locus, consisting of direct tandem repeats of a ~30 kb block containing four different genes, none of which resemble previously identified resistance genes (Cook et al., 2012; Cook et al., 2014; Lee et al., 2015). Compared to SCN-susceptible soybeans, only 1 of the 4 *Rhg1* repeat genes - an  $\alpha$ -SNAP, a vesicular trafficking chaperone – contains amino acid polymorphisms. Recently it was shown that *Rhg1*  $\alpha$ -SNAPs are impaired in normal interactions

with NSF, a key housekeeping ATPase, which together with  $\alpha$ -SNAP, maintains vesicular trafficking through disassembling SNARE trafficking protein complexes (Bayless et al., 2016).

Among soybean breeders, two phenotypic classes of *Rhg1* resistance derived from PI 88788 and PI 548402 have long been known (Niblack et al., 2002; Brucker et al., 2005). It is now known that two sources differ by *Rhg1* copy number and encode similar, but unique  $\alpha$ -SNAPs which are impaired in normal NSF-interactions (Cook et al., 2014; Lee et al., 2015; Bayless et al., 2016). The so-called high copy *Rhg1* loci contain up to 10 tandem repeats and are near exclusively used by soybean breeders to confer resistance in commercial soybeans (Brucker et al., 2005; Niblack et al., 2008). However, continued agricultural over-use of this high copy *Rhg1* from PI 88788 is selecting for SCN populations with increased virulence against this source (Niblack et al., 2008). Additionally, the low copy *Rhg1* loci requires an unlinked *Rhg4* locus on a different chromosome for effective SCN resistance (Meksem et al., 2001; Liu et al., 2012). However, because low copy *Rhg1* loci still confer resistance to some SCN populations virulent on high-copy *Rhg1* varieties, there is considerable agronomic interest in utilizing this alternate *Rhg1* resistance source (Brucker et al., 2005; Liu et al., 2012).

In this study, we examined the native genomic low copy *Rhg1* locus and uncovered an intact 4.77 kb Copia family LTR retrotransposon within  $\alpha$ -SNAP<sub>*Rhg1*</sub>LC intron 1. Notably, this Copia insertion is absent from examined single copy *Rhg1* (SCN-susceptible) and high copy *Rhg1* varieties. Hence, this Copia integration event occurred after *Rhg1* haplotype divergence. Both high and low copy *Rhg1* haplotypes were previously examined in next generation sequencing studies, however, this Copia insertion was likely tossed out during assembly (Cook et al., 2014; Liu et al., 2017). We investigated epigenetic as well as direct effects of this Copia element on  $\alpha$ -SNAP<sub>*Rhg1*</sub>LC protein production. However, whether this  $\alpha$ -SNAP<sub>*Rhg1*</sub>LC Copia

retrotransposon is domesticated and has regulatory functions at low copy *Rhg1* loci was indeterminate. Nonetheless, this study uncovers and draws attention to a potentially domesticated retrotransposon within a key resistance gene of certain haplotypes of the agriculturally valuable *Rhg1* resistance locus.

### 6.3 Results

#### **$\alpha$ -SNAP encoded by *Rhg1* low copy (*Rhg1<sub>LC</sub>*) haplotypes harbors an intronic *Copia* retrotransposon**

SCN-resistant soybeans used on farms almost exclusively utilize *Rhg1* high copy (*Rhg1<sub>HC</sub>*) resistance and SCN populations with increased virulence against *Rhg1<sub>HC</sub>* are emerging and spreading (Concibido et al., 2004; Niblack et al., 2008; Cook et al., 2012; Mitchum, 2016). However, low copy *Rhg1* (*Rhg1<sub>LC</sub>*) haplotypes, which possess a different  $\alpha$ -SNAP allele and have fewer *Rhg1* repeats, are reported to be efficacious against some *Rhg1<sub>HC</sub>*-selected SCN populations (Niblack et al., 2002; Brucker et al., 2005; Liu et al., 2012; Cook et al., 2014). For studies to assess *Rhg1<sub>LC</sub>* efficacy in transgenic soybeans, we amplified and sub-cloned the native *Rhg1<sub>LC</sub>*  $\alpha$ -SNAP locus from PI 89772 (*Rhg1<sub>LC</sub>*). Unexpectedly, the native low copy  $\alpha$ -SNAP PCR amplicon we obtained was approximately 5 kb larger than predicted from previous Whole Genome Sequencing (WGS) studies on *Rhg1* loci (Fig. S1A) (Cook et al., 2014; Liu et al., 2017).

Sanger sequencing of this unpredicted *Rhg1<sub>LC</sub>*  $\alpha$ -SNAP amplicon matched the predicted PI 89772 sequence until mid-intron 1, where a 4.77 kb insertion - not previously recognized as part of the *Rhg1* locus - was present (Fig. 1A). Nucleotide BLAST searches indicated that this insert was a member of the Ty-1 *Copia* superfamily of LTR retrotransposons. We subsequently named this insert “RAC”, for *Rhg1<sub>LC</sub>*  $\alpha$ -SNAP *Copia* (RAC) (Fig. 1A for *Rhg1<sub>LC</sub>*  $\alpha$ -SNAP-RAC model). Additionally, sequencing indicated that RAC was a fully intact retroelement, having both 5' and 3' LTRs (Long Terminal Repeats) and an uninterrupted ORF encoding a predicted 1,438 residue polyprotein. See Fig. S2 for complete RAC nucleotide sequence with flanking *Rhg1<sub>LC</sub>*  $\alpha$ -SNAP exons and RAC polyprotein translation. Analysis of the RAC polyprotein sequence indicated that conserved retrotransposon functional motifs for GAG, Protease,

Integrase and Reverse Transcriptase were present, suggesting that *RAC* could be a potentially active retrotransposon. (Kanazawa et al., 2009). Intriguingly, the LTR promoter and ORF of *RAC* were positioned anti-sense to the  $\alpha$ -SNAP ORF and promoter, with the *RAC* insert being just 396 bp away from the  $\alpha$ -SNAP ATG. The *RhgI<sub>LC</sub>*  $\alpha$ -SNAP amplicon sequence following the 4.77 kb *RAC* insertion matched WGS predictions.

***RAC* is integrated within  $\alpha$ -SNAP from all examined *RhgI<sub>LC</sub>* varieties, but is absent from the *RhgI<sub>HC</sub>* and SCN-susceptible Williams82 (Wm82) encoded  $\alpha$ -SNAP**

LTR retrotransposons may integrate directly within or adjacent to host genes (Havecker et al., 2004; Schulman, 2013; Galindo-Gonzalez et al., 2017). The *RAC* integration within *RhgI<sub>LC</sub>*  $\alpha$ -SNAP creates unique 5' and 3' sequence junctions within  $\alpha$ -SNAP<sub>LC</sub> intron 1. If *RAC* was unique to the PI 89772 *RhgI<sub>LC</sub>*  $\alpha$ -SNAP locus, or present in other *RhgI<sub>LC</sub>* haplotypes, as well as *RhgI<sub>HC</sub>* haplotypes, was unclear. Therefore, we screened genomic DNAs of the *RhgI<sub>LC</sub>* and *RhgI<sub>HC</sub>* accessions commonly used in HG type tests for the unique 5' and 3' intron 1-*RAC* junctions, using primers flanking the integration sites (Niblack et al., 2002). Agarose gel electrophoresis revealed both 5' and 3' intron 1-*RAC* junction products were detected in all *RhgI<sub>LC</sub>* varieties ("Forrest", PI 90763, PI 89772, PI 437654), but no *RAC* junctions were detected in the *RhgI<sub>HC</sub>* haplotype accessions (PI 88788, PI 548316, PI 209332) (Fig. 1B). The Williams82 (Wm82, SCN-susceptible) soybean reference genome sequence does not indicate a *RAC* insertion within the  $\alpha$ -SNAP locus, and we confirmed this absence using the same PCR assay on Wm82 gDNA (Fig. S1B). However, since *RhgI<sub>LC</sub>* haplotypes carry up to 3 *RhgI* repeat blocks, it was unknown if *RAC* was present in all repeat blocks vs. just 1 or 2. As such, we performed PCR screens similar to Fig. 1B, but with a primer pair flanking outside the *RAC*

integration, which would only produce a small amplicon if *RAC* is not inserted within  $\alpha$ -SNAP intron 1. Using the same gDNA stocks as in Fig. 1B, no wild-type intron 1 products were observed from PI 89772, PI 437654 and PI 90763, while the intron 1 product was present in Wms82 and in all *RhgI*<sub>HC</sub> haplotype varieties. (Fig. 1C). Together, these results indicate that *RAC* is similarly integrated within the  $\alpha$ -SNAP encoded by all *RhgI*<sub>LC</sub> repeats, but not in the  $\alpha$ -SNAP of *RhgI*<sub>HC</sub> or Wm82 (SCN-susceptible). That *RAC* is found only within *RhgI*<sub>LC</sub> haplotypes suggests that *RAC* integration was an event that occurred after the diversification of multi-copy *RhgI* haplotypes.

### **Copia elements similar to *RAC* are not abundant in Wm82 reference genome**

Copia (for copious) retrotransposons were named so because of their often high abundance (copy numbers) in plant or animal genomes (Du et al., 2010; Du et al., 2010; Zhao and Ma, 2013). If the *RAC* Copia family was abundant in the *Glycine max* genome, was unclear, however. We performed a BLAST of *RAC* against the soybean transposon database SoyTE, which contains over 32,000 transposons, including nearly 5,000 intact retrotransposons (Du et al., 2010). However, no significant hits were returned. We therefore examined the Wm82 soybean reference genome for *RAC* like elements using similar BLAST searches at Phytozome.org (Goodstein et al., 2012). Nucleotide BLAST produced just several intact and highly similar sequences to *RAC* - on Chromosomes 10, 09, 20 and 18, among others. Curiously, an intronic integration similar to that within *RhgI*<sub>LC</sub>  $\alpha$ -SNAP was also present within *Glyma.18G268000.1*, a putative leucine-rich repeat receptor kinase. Hundreds of small fragments, presumably LTR fragments with high identity to *RAC* were abundant across all *G. max* chromosomes (not shown). A NCBI BLAST of the predicted *RAC* polyprotein produced

several ~60% identity matches within pigeon pea (*Cajanus cajan*), *Glycine soja* and clover (*Trifolium subterraneum*), suggesting that this particular Ty-1 Copia family is at least common in legumes (not shown)

**The  $\alpha$ -SNAP<sub>*Rhg1*LC</sub>-RAC junctions are methylated, but RAC transcripts are readily detectable in *Rhg1*<sub>LC</sub> variety Forrest**

Typically, cells safeguard against transposons insertions by silence repetitive elements using small RNA directed DNA methylation (Kim and Zilberman, 2014). Because  $\alpha$ -SNAP<sub>*Rhg1*LC</sub> protein is readily detectable by Western blots in both root and leaf tissues, *RAC* apparently does not interrupt  $\alpha$ -SNAP<sub>*Rhg1*LC</sub> expression (Bayless et al., 2016). Regardless, we examined the  $\alpha$ -SNAP-*RAC* junction for DNA methylation using the restriction enzyme McrBC, which only cleaves methylated DNA. Briefly, after McrBC digestion, methylated genomic DNA regions can be interrogated by an attempted PCR over the region of interest. Comparing McrBC treated vs. mock treated genomic DNA of *Rhg1*<sub>LC</sub> varieties Forrest and Peking, we detected methylation across both 5' and 3' *RAC* integration sites (Fig. 2A).

It was unclear if methylation could be silencing *RAC* transcripts. Since *RAC* contained features of autonomous retrotransposons (intact LTRs, continuous ORF), we examined the transcript abundance of *RAC* (and similar cross-amplifying Copia elements), relative to a single copy locus adjacent to *Rhg1* (*Glyma18g02570*). qPCR indicated that *RAC* transcripts were ~200-fold higher in Forrest(*Rhg1*<sub>LC</sub>), as compared with Wm82 or Fayette(*Rhg1*<sub>HC</sub>), in roots or leaves (Fig. 2B). However, if these observed *RAC* transcripts were originating from the *Rhg1*<sub>LC</sub>  $\alpha$ -SNAP locus vs. another similar endogenous Copia element of Forrest was unclear. To distinguish between transcripts from *RAC* vs. other elements, we generated a native *Rhg1*<sub>LC</sub>  $\alpha$ -

SNAP locus with a unique nucleotide tag in the *RAC* ORF, and transformed this construct into Wm82 (Fig. 2C). Because sharp expression contrasts were noted by qPCR, we used RT-PCR to examine potential *RAC* expression. As in Fig. 2B, clear differences in *RAC* abundance from Wm82 vs. Forrest were observed using a primer amplifying the native *RAC*. Additionally, the unique *RAC* tag primer set indicated that transgenic addition of the native *Rhg1<sub>LC</sub>*  $\alpha$ -SNAP locus (with *RAC*) into Wm82 substantially increased *RAC* transcripts, suggesting that transcripts detected by the WT *RAC* primer likely derive from the *Rhg1<sub>LC</sub>*  $\alpha$ -SNAP-*RAC* integration (Fig. 2C). *Glyma18g02570* transcript amplification served as a cDNA control (Fig. 2C). Together, these results suggest that while methylation at *Rhg1<sub>LC</sub>*  $\alpha$ -SNAP-*RAC* junctions is detectable, transcripts with high identity to *RAC* are readily detectable in Forrest (*Rhg1<sub>LC</sub>* variety), but not Fayette (*Rhg1<sub>HC</sub>*) or Wm82 (single copy *Rhg1*), and likely derive from the native *Rhg1<sub>LC</sub>*  $\alpha$ -SNAP locus (with *RAC* insertion) vs. other endogenous soybean *Copia* elements.

### **Heat stress reduces *Rhg1<sub>LC</sub>* $\alpha$ -SNAP expression**

High levels of  $\alpha$ -SNAP<sub>*Rhg1<sub>LC</sub>*</sub> inhibited vesicle trafficking and were cytotoxic to *N. benthamiana* (Bayless et al., 2016). While protective host cell mechanisms silence transposons, certain stresses are known activate particular LTR retrotransposons (Woodrow et al., 2010; Matsunaga et al., 2012; Cavrak et al., 2014). Likewise, we examined if  $\alpha$ -SNAP<sub>*Rhg1<sub>LC</sub>*</sub> expression was affected by stresses, such as heat stress. We placed Forrest(*Rhg1<sub>LC</sub>*) roots at room temperature or 37°C for 24 or 48 hrs and examined  $\alpha$ -SNAP<sub>*Rhg1<sub>LC</sub>*</sub> expression as compared with WT  $\alpha$ -SNAPs or NSF using immunoblots with previously described antibodies (Bayless et al., 2016). We observed that heat stress for 1 or 2 days reduced  $\alpha$ -SNAP<sub>*Rhg1<sub>LC</sub>*</sub> protein abundance,

however, we noted that the abundance of WT  $\alpha$ -SNAPs, as well as that of the NSF co-chaperone, were also diminished (Fig. 3).

### **Silencing *RAC* does not appear to affect *Rhg1<sub>LC</sub>* $\alpha$ -SNAP expression**

*RAC* insertion into  $\alpha$ -SNAP<sub>LC</sub> intron 1 is not deleterious as the 4.77 kb element is spliced from the mature  $\alpha$ -SNAP<sub>LC</sub> mRNA and  $\alpha$ -SNAP<sub>Rhg1LC</sub> protein is readily detectable by immunoblot (Cook et al., 2014; Bayless et al., 2016). Nonetheless, it was unclear if abundant anti-sense RNAs targeting *RAC* could influence  $\alpha$ -SNAP<sub>LC</sub> expression, potentially through destabilizing the un-processed pre-mRNA and/or by effecting epigenetic changes at *RAC* (and thus within  $\alpha$ -SNAP<sub>Rhg1LC</sub>). To test this hypothesis, we generated a short hairpin RNA (shRNA) construct targeting ~250 bp of the *RAC* 3' ORF, and assessed  $\alpha$ -SNAP<sub>Rhg1LC</sub> expression using immunoblots in Forrest (*Rhg1<sub>LC</sub>*) hair roots silenced for *RAC* vs. an empty vector shRNA. shRNAs targeting *RAC* were not observed to consistently alter  $\alpha$ -SNAP<sub>Rhg1LC</sub> protein expression compared to empty vector silencing controls (Fig. 4A). As an alternate approach, we examined if mimicking *RAC* expression might elicit host silencing of  $\alpha$ -SNAP<sub>Rhg1LC</sub>. Therefore, we cloned the *RAC* ORF under the constitutive soybean ubiquitin promoter (*GmUbi*) and generated transgenic hairy roots of Forrest, as in Fig. 4A. We determined that constitutive expression of the *RAC* ORF did not impact  $\alpha$ -SNAP<sub>Rhg1LC</sub> protein abundance compared to empty vector controls (Fig. 4B).

### **Removing *RAC* from *Rhg1<sub>LC</sub>* $\alpha$ -SNAP does not alter *Rhg1<sub>LC</sub>* $\alpha$ -SNAP expression or splicing**

What impacts, if any, that *RAC* might have on  $\alpha$ -SNAP<sub>LC</sub> were not obvious. We directly examined if *RAC* impacted constitutive expression of  $\alpha$ -SNAP<sub>Rhg1LC</sub> protein by seamlessly deleting the 4.77 kb *RAC* integration within the native  $\alpha$ -SNAP<sub>LC</sub> construct using PIPE (Klock

and Lesley, 2009). We then transformed the native  $\alpha$ -SNAP<sub>LC</sub> or no *RAC* version into Wm82 and examine  $\alpha$ -SNAP<sub>Rhg1LC</sub> expression with immunoblots, as in Fig. 4A. No consistent differences in  $\alpha$ -SNAP<sub>LC</sub> expression were observed among Wm82 roots transformed with the native  $\alpha$ -SNAP<sub>LC</sub> (contains *RAC*) as compared with the otherwise native no *RAC*  $\alpha$ -SNAP<sub>LC</sub> (Fig. 5A). Previous reports showed that the  $\alpha$ -SNAP<sub>LC</sub> mRNA transcript is alternatively spliced (Cook et al., 2014; Bayless et al., 2016). This alternate splicing is likely caused by a C to G SNP within exon 6 that creates a putative GT splice donor (Blencowe, 2006; Cook et al., 2014). Because transposons inserted close or within to host genes have been reported to influence RNA splicing, we examined if the production of the previously reported  $\alpha$ -SNAP<sub>LC</sub> splice isoform was affected by the presence or absence of *RAC* (Krom et al., 2008). We transformed Wm82 with the native  $\alpha$ -SNAP<sub>LC</sub> or native no *RAC*  $\alpha$ -SNAP<sub>LC</sub> and amplified cDNA with a primer pair flanking the  $\alpha$ -SNAP<sub>LC</sub> alternative splice site. Using agarose gel electrophoresis, we examined the abundance of either splice isoform in the presence or absence of *RAC* - no effects on  $\alpha$ -SNAP<sub>LC</sub> splicing were apparent (Fig. 5B).

## 6.4 Discussion

*Rhg1* resistance-type  $\alpha$ -SNAPs play a pivotal role in the *Rhg1*-mediated resistance response to SCN (Bayless et al., 2016). Our study that an intact Copia element (*RAC*) lies within the  $\alpha$ -SNAP encoded by repeats of *Rhg1* low copy haplotypes deepens our understanding of both *Rhg1* structure and divergence. That *RAC* is not inserted within the  $\alpha$ -SNAP encoded by *Rhg1*<sub>HC</sub> or by SCN susceptible soybeans suggests that *RAC* integration occurred after the development of the two multi-copy *Rhg1* classes. Being integrated within intron-1 and anti-sense to  $\alpha$ -SNAP<sub>LC</sub>, *RAC* is suitably positioned to modulate  $\alpha$ -SNAP expression, yet without directly interrupting the  $\alpha$ -SNAP<sub>LC</sub> ORF. Although we observed methylation near the  $\alpha$ -SNAP<sub>LC</sub>-*RAC* integration site, *RAC* transcripts originating from the  $\alpha$ -SNAP<sub>LC</sub>-*RAC* locus were readily detectable. However, we did not detect any impacts of *RAC*, or of small RNAs targeting *RAC*, on normal  $\alpha$ -SNAP<sub>LC</sub> expression. Nevertheless, this study highlights that further differences exist amongst the two resistance-conferring *Rhg1* haplotypes, and indicates that *Rhg1*<sub>LC</sub> repeat blocks contain an additional 5kb gene. Additionally, this study underlines the potential drawbacks of using short-read WGS to assess allelic differences, as reads for the  $\alpha$ -SNAP<sub>LC</sub>-*RAC* region were apparently tossed during genome assembly in previous studies due to similarity to other genomic reads (Cook et al., 2014; Liu et al., 2017).

The observation that *RAC* is integrated within all examined *Rhg1*<sub>LC</sub> haplotypes supports that *RAC* could confer a beneficial trait and was perhaps selected and maintained within *Rhg1*<sub>LC</sub> haplotypes.  $\alpha$ -SNAPs are essential eukaryotic housekeeping genes that maintain vesicle trafficking, and wild-type (WT)  $\alpha$ -SNAP expression is sharply diminished in *Rhg1*<sub>LC</sub> haplotypes compared to either *Rhg1*<sub>HC</sub> or SCN susceptible single-copy *Rhg1* soybeans (Zhao et al., 2015)(Chapter V). Because high doses of  $\alpha$ -SNAP<sub>*Rhg1*LC</sub> impaired vesicular trafficking and were

cytotoxic to *N. benthamiana* while WT  $\alpha$ -SNAP co-expression alleviated this toxicity (Bayless et al., 2016), fine tuning  $\alpha$ -SNAP<sub>Rhg1<sub>LC</sub></sub> expression at particular developmental stages or environmental situations could be advantageous to *Rhg1<sub>LC</sub>* haplotypes. It is furthermore intriguing that the ORF for the RAC encoded polyprotein, at least from PI 89772, was intact with no stop codons, suggesting that *RAC* is an autonomous element (Schulman, 2013). While *RAC*- $\alpha$ -SNAP integration was detected in *Rhg1<sub>LC</sub>* haplotypes, whether SNPs or alterations are present within *RAC* are carried by different *Rhg1<sub>LC</sub>* varieties, is unclear. Future studies could examine *RAC* divergence within various *Rhg1<sub>LC</sub>* sources if SNPs are present.

Only a handful of intact *Copia* elements similar to *RAC* were identified by BLAST in the Wm82 reference genome and no similar elements were present in the Soy TE database (Du et al., 2010). This suggests that the *RAC* family in *Glycine max* is low copy and perhaps recent. However, that two of these *RAC*-like elements were positioned either intronically or near genes with putative roles in defense or development, was intriguing. Reports suggest that host silencing of transposons can establish *cis*-regulatory networks whereupon the expression of nearby host genes is also affected (Lisch and Bennetzen, 2011; McCue et al., 2012; McCue and Slotkin, 2012). If *RAC*, and other endogenous retrotransposons potentially underlie regulatory networks operating at certain development stages or environmental stresses is an intriguingly speculation that deserves further investigation.

If *RAC* influences the  $\alpha$ -SNAP<sub>LC</sub> epigenetic landscape beyond DNA methylation at the integration junctions (i.e., histone post-translation modifications) is unclear. Importantly, if *RAC* can modulate epigenetic marks in response to specific abiotic or biotic stresses is of interest (Slotkin and Martienssen, 2007; McCue et al., 2012; McCue et al., 2013). A recent study reported that cyst nematode infection of *Arabidopsis* causes hypomethylation of transposable

elements, including retrotransposons (Hewezi et al., 2017). If *RAC* activity is modified SCN syncytium formation and in turn, alters the *Rhg1<sub>LC</sub>* resistance response is an obvious area to further explore. Likewise, assessing *RAC* sensitivity to additional environmental stimuli and if these could affect  $\alpha$ -SNAP<sub>LC</sub> production, is of interest and could also be useful to the agricultural community. For example, if *Rhg1<sub>LC</sub>* soybeans are deployed to growing regions subject to stresses that affect *RAC* activity, SCN resistance could be impacted. Links between yield depression and *Rhg1<sub>LC</sub>* resistance are reported, however, the underlying mechanisms for this remain unclear. Whether  $\alpha$ -SNAP<sub>LC</sub> expression, and also *RAC* impact this observation is unclear.

The spread of SCN populations virulent against the standard *Rhg1<sub>HC</sub>* resistance is a concern to growers (Brucker et al., 2005; Niblack et al., 2008). *Rhg1<sub>LC</sub>* sources of resistance are a potential solution to maintain field SCN-resistance against these SCN. Although we did not detect influence of *RAC* on constitutive  $\alpha$ -SNAP<sub>LC</sub> expression, it remains to be determined if *RAC* is merely “hitching a ride” within a valuable host gene under positive selection, or if *RAC* could benefit the host by acting as a regulatory “domesticated transposon” (Tsuchiya and Eulgem, 2013). It is possible that *RAC* insertion is simply neutral and confers no advantage to *Rhg1<sub>LC</sub>* haplotypes. Nonetheless, identifying that *RAC* is present within *Rhg1<sub>LC</sub>* haplotypes expands our knowledge of the structure and divergence of the valuable *Rhg1* SCN resistance locus.

## 6.5 Materials and Methods

### DNA extraction

Soybean DNA was extracted from young soybean trifoliate or soybean roots using previously CTAB methods (Cook et al., 2012).

### RAC cloning

Approximately 100 ng of high quality CTAB gDNA was amplified with Kapa HiFi polymerase annealing at ~70°C for 30 seconds and extension for 5 minutes at 72°C with at least 32 cycles of amplification (KAPA Biosystems). The PCR fragment containing  $\alpha$ -SNAP<sub>Rhg1</sub>LC-RAC from PI 89772 was TA cloned using the Topo xL cloning kit (Life Technologies Corp., Carlsbad CA). For native  $\alpha$ -SNAP<sub>Rhg1</sub>LC expression assays, the ORF was cloned using AvrII – SbfI restriction sites with compatible XbaI PstI sites in the previously described pSM101 binary vector (Cook et al., 2012; Bayless et al., 2016).

### Methylation Analysis

Methylation studies on genomic DNAs of Forrest and Peking were performed similar to (Cook et al., 2014). Briefly, McrBC digests DNA with methylated cytosines in a sequence independent manner and does not cut unmethylated DNA. Methylation controls reactions contain the same amounts of DNA in reaction buffer, but with no McrBC enzyme added. McrBC digestion for at 37C for 90 minutes, and heat inactivated at 65C for 20 minutes. Samples were subsequently amplified via PCR and visualized by agarose gel electrophoresis.

### RNA isolation and cDNA synthesis

RNA was extracted using standard Trizol methods. Williams82, Fayette or Forrest cDNAs from root or leaf tissue were generated using the iScript cDNA Synthesis Kit (Bio-Rad).

### **qPCR analysis**

qPCR was performed on a CFX96 real-time PCR detection system (BioRad, Hercules, CA). RNAs were extracted with Trizol via standard methodologies or the Direct-Zol RNA kit according to manufacturer's recommendations (Zymo research, Irvine, CA). All RNA samples were quantified with a Nanodrop spectrophotometer and equal RNA inputs were added to each respective cDNA reaction. RNA quality was also assessed by running briefly on a 1.2% agarose gel in TBE-buffer. SsoFast EvaGreen Supermix (Biorad, Hercules, CA) was used for amplification detection and primer concentrations were at 0.2  $\mu$ M and 0.3  $\mu$ M. Two technical replicates were used per each sample. Following amplification, a melt curve program was performed.

### **Vector construction**

The *RAC* ORF was placed directly under the control of the soybean ubiquitin promoter in the vector pBlueScript using the polymerase incomplete primer extension (PIPE) method (Klock and Lesley, 2009) and sequence verified. Binary vectors containing native  $\alpha$ -SNAP<sub>Rhg1</sub>LC-*RAC* were generated via PCR amplification of  $\alpha$ -SNAP<sub>Rhg1</sub>LC-*RAC* with *AvrII* and *SbfI* ends and ligated into pSM101 binary expression cassettes were digested with *XbaI*/*PstI* (New England Biolabs). Gel extractions performed using the Qiaquick gel extraction kit (Qiagen) or the Zymoclean Large Fragment DNA Recovery Kit. Purified DNA fragments were then ligated into the binary vector pSM101 using T4 DNA ligase (New England Biolabs). Cloning into the

shRNAi2 vectors to generate the shRNA construct targeting RAC was performed as previously described (Cook et al., 2012). Briefly, *RAC* PCR products with restriction overhangs were generated and digested with AvrII-AscI and SwaI-BamHI restriction overhangs and cloned step-wise into pGRNAi2. Polymerase Incomplete Primer Extension (PIPE) was used to delete *RAC* from within the  $\alpha$ -SNAP<sub>*Rhg1*</sub>LC-RAC Topo xL subclone.

### **Transgenic soybean hairy root generation**

Binary expression constructs were transformed into *Agrobacterium rhizogenes* strain, “Arqua1”. All transgenic soybean roots were produced transforming soybean cotyledons with either pSM101 or pG2RNAi2 as described in (Cook et al., 2012).

### **Immunoblotting & antibodies**

Affinity-purified polyclonal rabbit antibodies raised against  $\alpha$ -SNAP<sub>*Rhg1*</sub>LC, wild-type  $\alpha$ -SNAPs and NSF were previously generated and validated using recombinant proteins and transgenic lysates in Bayless 2016.

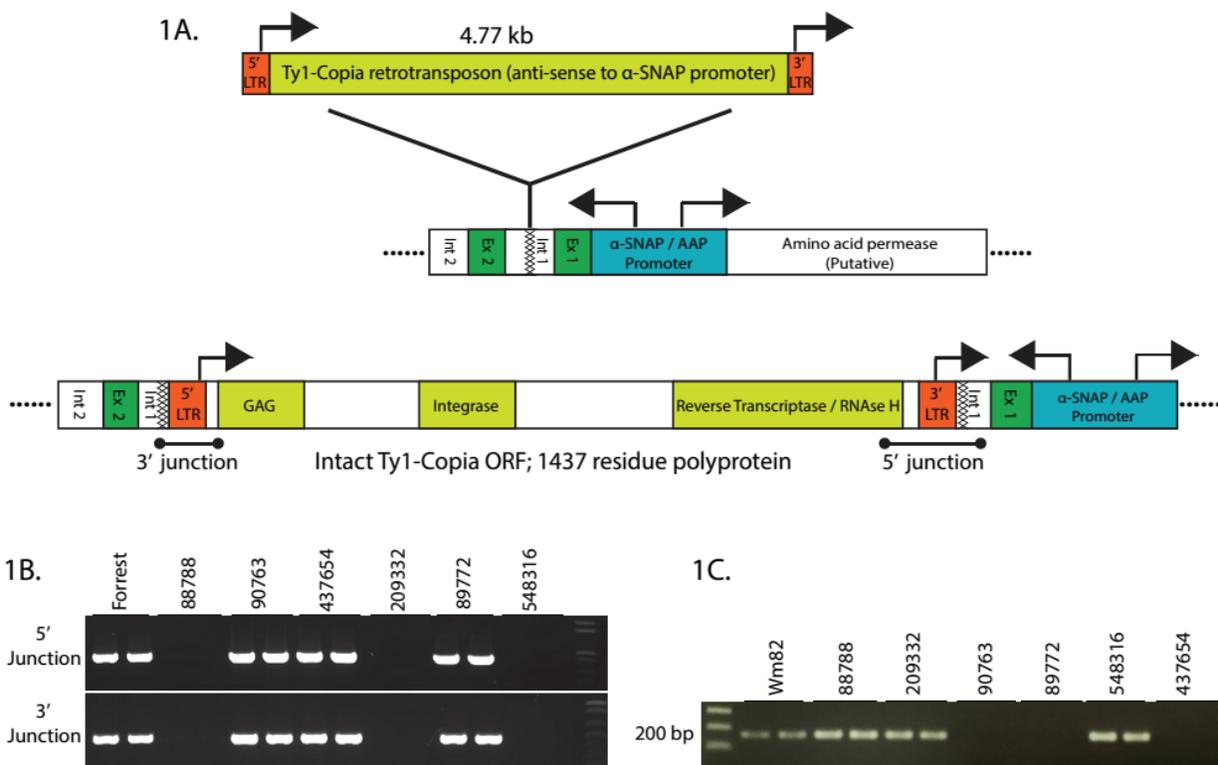
Tissue preparation and immunoblots were performed essentially as in (Song et al., 2015; Bayless et al., 2016). Soybean roots were flash-frozen in N<sub>2</sub>(L), massed, and homogenized in a PowerLyzer 24 (MO BIO) for three cycles of 15 seconds, with flash-freezing in-between each cycle. Protein extraction buffer [50 mM Tris·HCl (pH 7.5), 150 mM NaCl, 5 mM EDTA, 0.2% Triton X-100, 10% (vol/vol) glycerol, 1/100 Sigma protease inhibitor cocktail] was then added at a 3:1 volume to mass ratio and samples were centrifuged and stored on ice. Immunoblots for either *Rhg1*  $\alpha$ -SNAP were incubated overnight at 4 °C in 5% (wt/vol) nonfat dry milk TBS-T (50 mM Tris, 150 mM NaCl, 0.05% Tween 20) at 1:1,000. NSF immunoblots were performed

similarly, except incubations were for 1 h at room temperature. Secondary horseradish peroxidase-conjugated goat anti-rabbit IgG was added at 1:10,000 and incubated for 1 h at room temperature on a platform shaker, followed by four washes with TBS-T. Chemiluminescence detection was performed with SuperSignal West Pico or Dura chemiluminescent substrate (Thermo Scientific) and developed using a ChemiDoc MP chemiluminescent imager (Bio-Rad).

## **6.6 Acknowledgements**

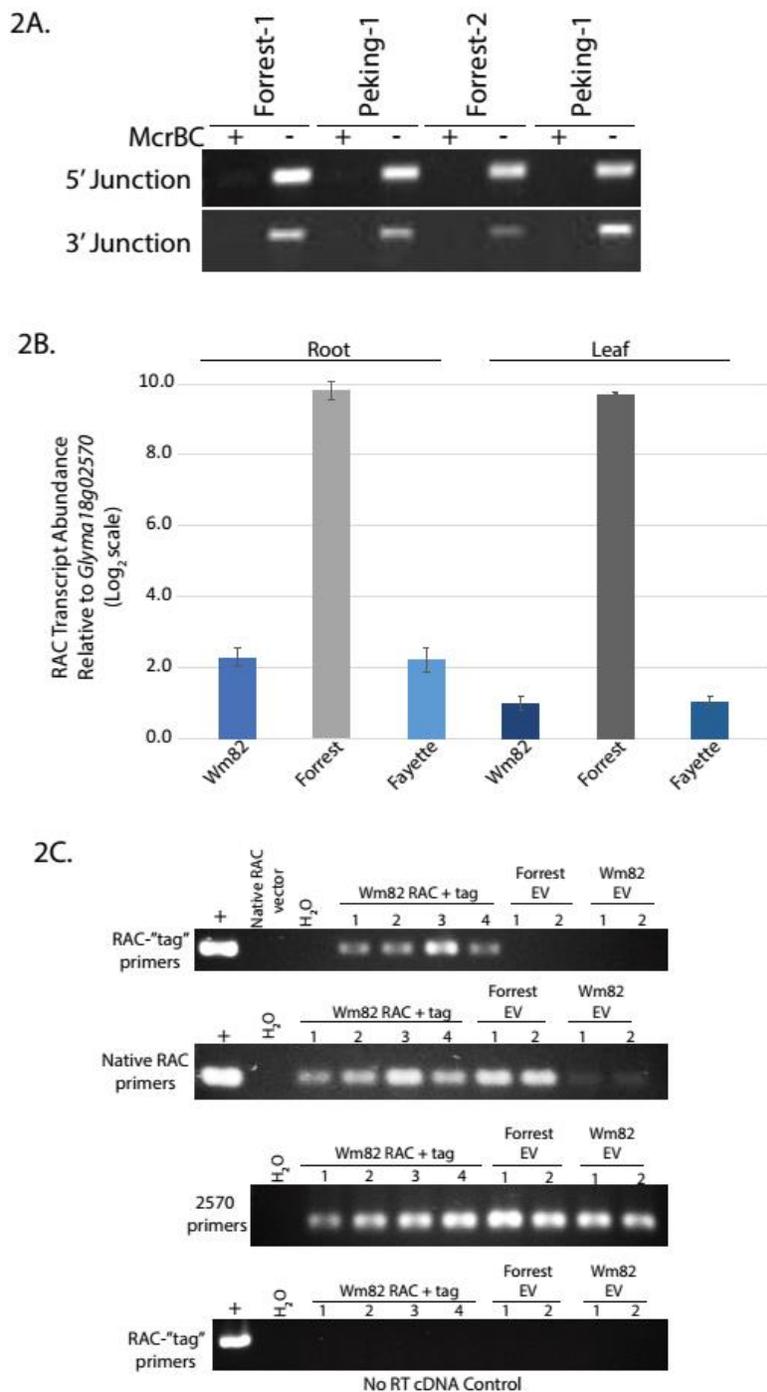
This work was funded primarily by USDA-NIFA-AFRI award 2014-67013-21775 to A.F.B., and also by the United Soybean Board. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. (DGE-1256259) to A.M.B.

## 6.7 Figures



**Figure 1.** Low copy haplotype *Rhg1* genomes contain an anti-sense Ty1-Copia family retrotransposon insertion at intron 1 of *Glyma.18g022500* ( $\alpha$ -SNAP).

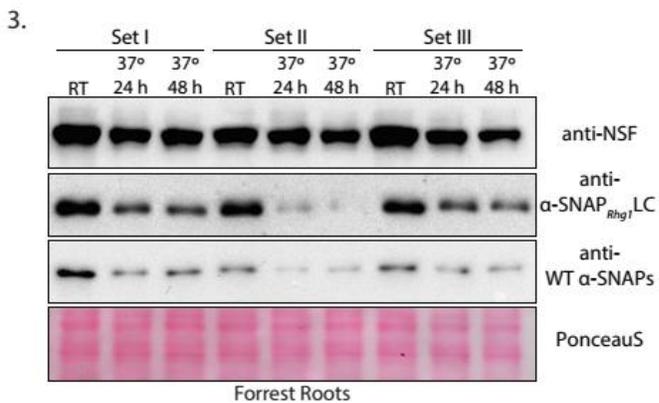
(A) Model of *Rhg1* associated Copia (RAC) integration in  $\alpha$ -SNAP<sub>*Rhg1*</sub>LC. 4.77 kb RAC is positioned anti-sense to  $\alpha$ -SNAP<sub>*Rhg1*</sub>LC ORF. RAC ORF uninterrupted and encodes putative 1437 residue polyprotein. RAC LTRs shown red, open reading frame in greenish yellow. LTR: long terminal repeat; GAG: group-specific antigen.  $\alpha$ -SNAP promoter and exons shown in blue or green, respectively. (B). Agarose gel showing 5' and 3'  $\alpha$ -SNAP-RAC amplicons from *Rhg1* low copy soybeans Forrest, 90763, 437654, 89772. Amplified junction regions indicated above on 1A model. No RAC junction amplicons were present for high copy type *Rhg1* lines PI88788, PI 209332, or PI548316. (C) Similar to B, but agarose gel showing PCR amplification of the uninterrupted  $\alpha$ -SNAP intron 1, as in the Wm82 soybean reference genome.



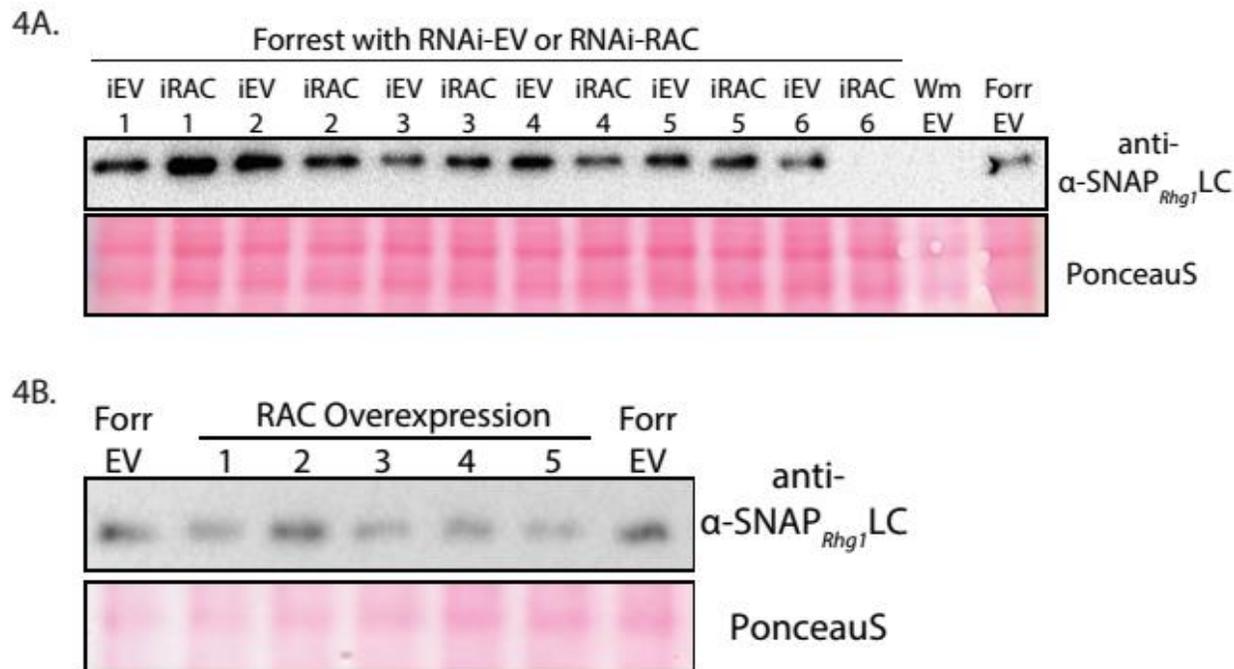
**Figure 2.** *RAC* integration junctions are methylated; *RAC*-like transcripts are relatively abundant in the low copy soybean, Forrest.

(A) Agarose gel of PCR amplicons spanning both 5' and 3'  $\alpha$ -SNAP-*RAC* junctions from soybean root genomic DNA treated with (+) or without (-) McrBC methylation sensitive

restriction enzyme (McrBC only cuts methylated DNA). (B). RAC (and similar element) transcript abundance in leaf or root cDNA of Wm82 (SCN-susceptible), Forrest (*Rhg1<sub>LC</sub>*) or Fayette (*Rhg1<sub>HC</sub>*), relative to Glyma18g02570 transcript abundance, using qPCR. Note: Y-axis presented in log<sub>2</sub> scale. (C) RT-PCR cDNA analysis of Wm82 roots transformed with native  $\alpha$ -SNAP<sub>LC</sub> (RAC + tag) or empty vector compared with Forrest empty vector, using the specified primer sets. The inserted RAC tag acts as a unique identifier of  $\alpha$ -SNAP<sub>LC</sub> RAC transcripts and does not detect endogenous elements of similar sequence, like the native *RAC* primer. Glyma1802570 transcript used as cDNA quality control. No RT (reverse transcriptase) control reactions indicate absence of gDNA contamination.

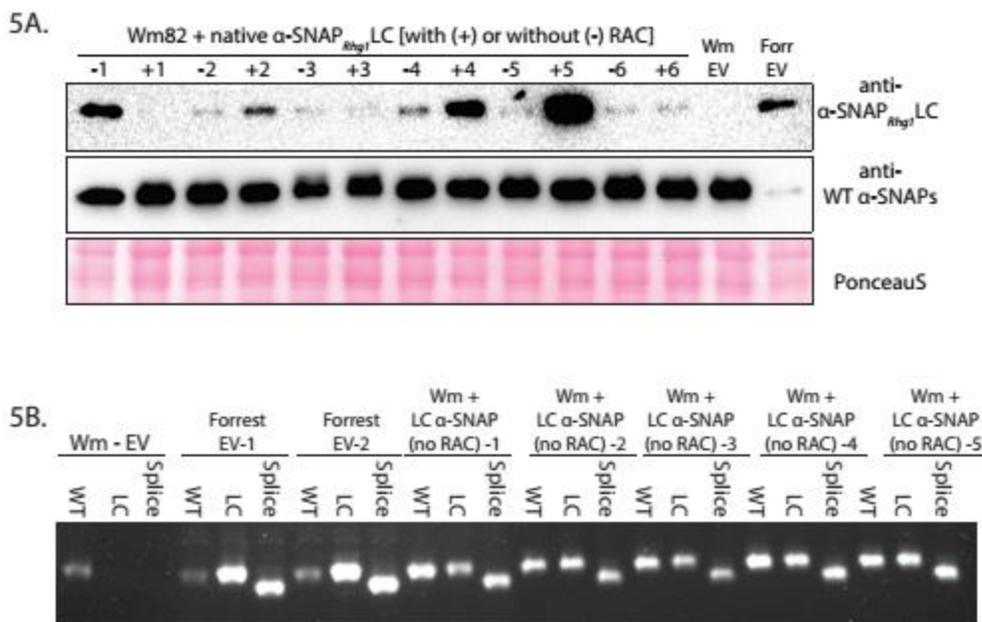


**Figure 3.** Heat stress treatment reduces  $\alpha$ -SNAP<sub>Rhg1</sub>LC and WT  $\alpha$ -SNAP protein abundance. Immunoblot of Forrest root lysates indicating  $\alpha$ -SNAP<sub>Rhg1</sub>LC, WT  $\alpha$ -SNAP or NSF protein expression after roots were placed in a 37°C heated incubator or a dark ambient cabinet for 24 or 48 hrs. PonceauS staining indicates total protein levels loaded per lane. Roots were grown and cultured in sterile environments on hairy root medium.



**Figure 4.** shRNA targeting of *RAC* does not appear to affect  $\alpha$ -SNAP<sub>Rhg1</sub>LC expression.

(A). Immunoblot indicating  $\alpha$ -SNAP<sub>Rhg1</sub>LC expression in Forrest roots transformed with shRNAi2-empty vector (iEV) or shRNAi2 constructs targeting *RAC* (iRAC). Williams82 (Wm, SCN-susceptible) and Forrest (Forr) empty vector roots included as negative and positive controls for  $\alpha$ -SNAP<sub>Rhg1</sub>LC expression, respectively. PonceauS staining indicates total protein amounts. (B). Immunoblot of  $\alpha$ -SNAP<sub>Rhg1</sub>LC, as in A, but from Forrest roots transformed with the *RAC* ORF constitutively expressed by the soybean ubiquitin promoter.

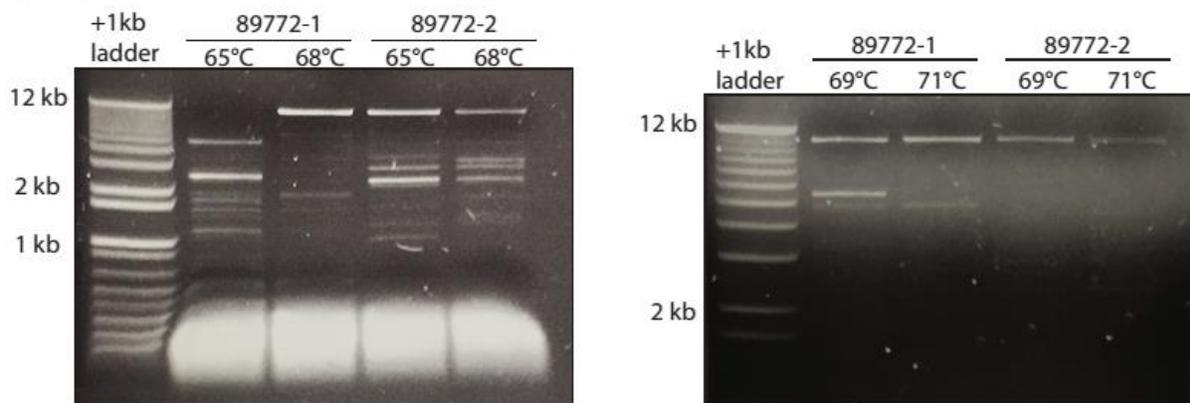


**Figure 5.** Removing *RAC* from  $\alpha$ -SNAP<sub>Rhg1</sub>LC may not alter protein expression or alternative splicing.

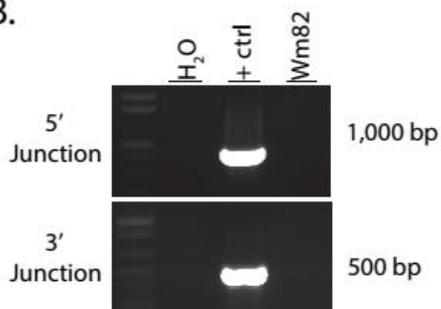
(A). Immunoblot of  $\alpha$ -SNAP<sub>Rhg1</sub>LC from Wm82 roots transformed with a native  $\alpha$ -SNAP<sub>Rhg1</sub>LC locus, with (+) or without (-) the *RAC* integration (no *RAC*). (B) PCR amplification of cDNA from Wm82 or Forrest roots transformed with empty vector, or of transformed with the same  $\alpha$ -SNAP<sub>Rhg1</sub>LC (no *RAC*) cassette. “WT” refers to amplification with a primer set identifying WT  $\alpha$ -SNAP transcripts, “LC” refers to primer set for full length  $\alpha$ -SNAP<sub>Rhg1</sub>LC transcript detection and “Splice” amplifies a previously identified  $\alpha$ -SNAP<sub>Rhg1</sub>LC isoform, which migrates at a smaller size.

## 6.8 Supplemental Figures

S1A.



S1B.



### Supplemental Fig 1A.

(A). Agarose gel showing PCR amplification of the native  $\alpha$ -SNAP<sub>Rhg1LC</sub> with RAC insertion from PI 89772 genomic DNA. Temperatures indicate the amplification annealing temperature with Kapa HiFi polymerase. (B). PCR amplification of the 5' and 3' RAC-  $\alpha$ -SNAP junctions using gDNA of Wm82. Positive PCR control is a subcloned RAC element from PI 89772 within a pTopo xL vector.

### Translation of *Rhg1<sub>LC</sub>* $\alpha$ -SNAP Copia (RAC) Polyprotein – 1438 residues

MESYLYLHPSENPAIALVSPVLDSTNYHSWSRSMVTALSAKNKVEFIDGSAPEPLKTDRMHGAW  
 RRCNNMVVSWIVHSVAISIRQSILWMDKAEIWRDLKSRYSQGDLLRISDLQQEASTMKQGTLT  
 VTEYFTRLRVIWDEIENFRPDPICSCNIRCSCNAFTIIAQRKLEDAMQFLRGLNEQYGNIRSHVLL  
 MDPIPAISKIFSVAQQRQLLGNAGPGIHFEPEKISINAAKTVCDFCGRVGHLESTCYKKGMP  
 NHDTRNKNSGRKACTHCGKMGHTVDVCYRKHGYPPGYTPGYKPYGGRTTVNNLVAVENKAN  
 EDQAQHHEAHSVRFSPSEQYKALLALIQEPSAGNTTISQSKQMASISSCINTPTNPGMSLSLRTPC  
 VSWILDSGATDHATYSLRNLHSYKQIDPITVKLPNGQVCATHSGIVKLSSNIILQDVLYIPSFTFN  
 IISIKLVSSVNCLEIFSSSTSCVLEQVNSHMRIGIVEAKHGLYHLIPAQLTTKTVNSTITHPRCNVIPI  
 DLWHFRLGHPSTERIQCMKAYYPLLKNNKDFVCNTCHHAKQKKLPFSLSHSHASHIFDLLHMDI  
 WGPCSKPSMHGHKYFLTIVDDCSRFTWVHLMKSKAETRHIIMNFITFIETQYDGVKVIIRSDNGIE  
 FSMHHYYASKGIIHQTTCIETPEQNGIVERKHQHLLNVTRALLFQASLPPSFWCYALPHATYLINE  
 IPTPFLHNVSPEYKHLKHPDISNLRVFGCLCYINTLKANRQKLDKAKAHPCIFIGFKMHTKGYLVY  
 DLHSNDVSIISRNITFYEDHFPYLSETQHTHLEYPAPSPESFSDRNDDPQTESLSSPPMISIPSSNEPEH  
 NHPPSHLRRSTRKNTPTYLRDYHREFASSTPSTMAVRYPLSSVLSYSRSLPAHRNFVMNISSVT  
 EPTSIVDASRHDCWIKAMEAELRALQSNQWTWRLTPLPSHKTAIGCRWVYKIKYRADGSIERHKA  
 RLVAAGYTMQEGLDYLDTFSPVAKLTTVRLLLAIAALNQWHLRQLDVNNAFLHGELDEEVYM  
 QIPGLSVDNPKLVCRLQKSLYGLKQASRQWFVKLSSFLTSHGFHQSTADHSLFLRFTGNITTILL  
 VYVDDIILTGNSMTEIQTIVTLLDSEFKIKDLGDLKFFLGLFIARSSKGIHLCQRKYTLDILNASGM  
 LGCKPNSTPIDYSTKLQADSGSPLSAESSSSYRRLIGKLIYLTNTRPDITYAVQQLSQYMAAPTNA  
 HLQAAFRLRYLKSSPGSGIFFTAAGTAQLRAFSDDWAGCKDSRKSTTGYL VYFGSSLVSWQSK  
 KQPTVSRSSSEAEYRALASTTCELQWLTFLQDFRISFVQPANLYCDNQSAIQIATNPVFHERTKH  
 IEIDCHIVRQKLNGLLKLKLPVSSALQLADIFTKALAPAVFRHLCKNLGMMNIHSQLEGG

### $\alpha$ -SNAP<sub>*Rhg1*</sub>LC : RAC Integration Sequence

#### Sequence key:

$\alpha$ -SNAP Exon 1

$\alpha$ -SNAP Intron 1

RAC 3' LTR

RAC Polyprotein ORF

RAC 5' LTR

$\alpha$ -SNAP Exon 2

ATGGCCGATCAGTTATCGAAGGGAGAGGAATTCGAGAAAAAGGCTGAGAAGAAGCTCAGC  
 GGTTGGGGCTTGTGGCTCCAAGTATGAAGATGCCGCCGATCTCTTCGATAAAGCCGCCAA  
 TTGCTTCAAGCTCGCCAAATCATGTTTTTCCTCTTCTCTCTACTTTTTTAAATTCCATTCGT  
 GTCTCCTCAAAATGTTGATTTAGTGTGCATAAATCATAATTATTATTCTCTTCTATTGTTGTTAT  
 TTTATTGTTATTACTTCAATCGACGAGTGTGTTGAGTTTTGAGGTGTCCGATTTCCCGATTAA  
 TTGAAGTATAGTTTTAATCTGATTTACTGGAAAATATTTTTTTGCCTGATTTTGTTTTTTGG  
 AACAATTACTAGCATATAAATTAGAATTGTGGATGAAGTA TAAGAAAACACTACAGGAGCATA  
 AGGAAGAAGAAGTGAGCTTGAATATTCAGAGAAGAAGATCAGCTTCAGCTACTTATTTTCG

TATTAACAGAGAAGGTTTATATACATGATGTGTGATTGTTATAACAGAAAAGCTAACTAACT  
 CAACTAACCCAACTACCCTTAACTGATACTGTTATACTGCTAAGA GCCCCCTCAAGCTGGG  
 AATGGATATTCATCATTCCCAGCTTGTTACAGAGATGCCGAAAAACAGCAGGTGCAAGAGCT  
 TTGGTGAATATATCCGCTAGTTGTAAAGCTGACGAAACCGGAAGAAGCTTTAGGAGACCTG  
 AGTTAAGCTTTTGACGAACAATATGACAGTCTATCTCGATATGTTTAGTTCGTTCGTGAAAA  
 ACGGGATTAGTAGCTATTTGGATGGCTGACTGATTATCACAGTATAAGTTCGCTGGTTGAAC  
 GAATGAGATACGAAAGTCTTGAAGCAAGAAGGTGAGCCACTGTAGTTCGCAAGTAGTGGAG  
 GCAAGAGCGCGGTATTCAGCTTCGGAAGAAGTTCGCTGACACAGTTGGCTGCTTCTTGGATTG  
 CCAAGAAACCAGAGAGGAACCAAAGTAACTAAGTAGCCGGTAGTGGATTTCTTGAATCT  
 TTGCATCCAGCCCAATCCGAGTCACTAAAGGCTCGGAGTTGTGCGGTACCTGCGGCAGTGAA  
 GAAGATACCTGATCCCGGAGAAGTCTTGAGGTATCGAAGAATCCGAAAGGCGGCTTGAAGA  
 TGAGCATTGGTGGGGCGGCCATGTACTGGCTGAGTTGTTGAACAGCATACTTATATCGGG  
 CCTGGTGTGGTAAGGTATATTAATTTACCGATCAATCGCCGATAAGAGGAAGAAGACTCAG  
 CTGAGAGAGGACTGCCGAATCTGCCTGTAAGTTCGTAGAGTAGTCTATTGGTGTGAATTG  
 GGCTTGCATCCCAGCATTCCGGATGCATTTAGAATGTCTAATGTATACTTGCCTGGCATAA  
 GTGTATCCCTTCGAGCTTCGGGCGATTCAAGCCCAAGGAAAAACTTTAAATCCCAAGAT  
 CCTTGATCTTAAATTCAGAATCTAAGAGGGTGACAATTGTTTGTATTTTCGGTTCATGCTATTT  
 CTGTGAGAATGATGTCGTCTACATAAACAAGAAGGATGGTTGTGATGTTTCCAGTAAACCGC  
 AAAAAGAGAGAGTGTATCCGAGTTGACTGATGAAAGCCATGAGAGGTTAAGAAGCTTGACA  
 ATTTTACGAACCATTGTCTGCTGGCTTGTGAGACCATAAAGAGACTTTTGAAGGCGACAT  
 ACAAGCTTTGGGTTATCAACGAAAGTCCCGGAGGTATTTGCATATAAACCTCCTCGTCAAG  
 TTCTCCATGGAGGAATGCATTATTAACATCTAGCTGCCGTAATGCCACTGGTTAATGCTGC  
 AATTGCAAGAAGAAGGCGCACCGTGGTTCAGCTTTGCTACCGGAGAGAAAGTGTCAAGGTAG  
 TCTAACCTTCCATTTGGGTGTATCCCTTTCGCAACCAGCCGCGCTTTATGCCTTTCGATGGAT  
 CCATCTGCTCTATACTTTATTTATAGACCCATCTGCACCCAATAGCCGCTTGTGAGAAGGG  
 AGAGGTGTGAGGCGCCATGTTTGGTTCGACTGAAGAGCTCGTAGCTCGGCTTCCATGGCCTT  
 AATCCAGCAATCATGGCGAGAAGCATCGACATATGAGGTTGGCTCTGTGACGGAGGAAATA  
 TTCATGACAAAGTTCCTGTGGGCAGGAGACAAGCGTGAGTAAGAGAGTACGGAACATAAGTG  
 GATAACGAACAGCCATTGAAGTGTCTGGTGTAGAGGAAGCAAAGTCTCTGTGGTAATCTCGG  
 AGGTACGTTGGGGTGTGTTTGGTCTGGTGGATCGTCTAAGGTGTGAAGGAGGATGATTATG  
 TTCAGGTTCAATTGATGATGGTATAGAGATCATAGGTGGTGTGAAAGACTCTCTGTTTGTG  
 GGCATCGTTTCTGTCGGAGAAGGATCCCGGTGAGGGAGCTGGATATTCTAAGTGTGTATGC  
 TGAGTTTCAGAAAGATAAGGAAAATGATCCTCATAAAATGTGATATTTTCGAGAGATGCTAAC  
 ATCATTAGAGTGCAAATCATAACACAAGATATCCCTTTGTATGCATTTAAAACCGATGAATA  
 TGCATGGATGAGCCTTAGCATCAAGCTTTGCCGTTTGCCTTGTAGTGTATTTATGTAACATA  
 GACACCCGAAAACACGAAGGTTAGAAAATGTCACAAGGGTGTATATGCAGCTTTTCATAGGGT  
 GAAACATTATGCAAAAACGGCGTGGGAATACAATTAATCAAGTAAGTGGCATGCGGCAAAG  
 CGTAACACCAGAAGCTTGGTGGTAGACTTGCCTGAAACAAAAGTGCACGTGTGACATTGAG  
 AAGGTGCTGGTGTGTTGCGTCTACAATTCGTTTGTCTGGAGTTTCAATGCACGTGGTCTG  
 GTGTATGATGCCCTTTGATGCATAGTAATGATGCATGGAGAATTCAATGCCATTATCACTTCT  
 GATGATCTTAACCTTGCATCGTATTGTGTTTCAATGAATGAATGAAGTTCATGATTATATG  
 TCGGGTTTCAGCTTTGGATTCATAAAGATGAACCCATGTAAAGCGTGAGCAATCATCAACTA  
 TAGTTAAGAAGTATTTGTGCCATGCATGGATGGTTAGAGCACGGACCCCATATGTCCATA  
 TGCAGTAAGTCAAAAATGTGAGATGCATGTGAATGGCTAAGAGAAAAAGGTAATTTCTTTG  
 TTTTCGATGATGACACGTGTGCAACAAAATCCTTATTATTTTGAAGAAGGGGATAGTAAG  
 CTTTCATACATTGATTCTTTTCAGTGGATGGGTGGCCTAACCTAAAATGCCAAAGGTCAATA  
 GGTATTACATTACATCGAGGGTGAGTAATAGTGGAGTTTACGGTTTTGGTGGTCAGCTGAGC

**Supplemental Fig. 2.** (A). Translation of the putative *RAC*-encoded polyprotein. (B). Sequence

of the *RAC* integration within  $\alpha$ -SNAP<sub>Rhg1</sub>LC of PI 89772.  $\alpha$ -SNAP<sub>Rhg1</sub>LC and *RAC* ORF sequences are color coded as indicated in the above key.

## 6.9 References

- Bayless AM, Smith JM, Song J, McMinn PH, Teillet A, August BK, Bent AF** (2016) Disease resistance through impairment of alpha-SNAP-NSF interaction and vesicular trafficking by soybean Rhg1. *Proc Natl Acad Sci U S A* **113**: E7375-E7382
- Blencowe BJ** (2006) Alternative splicing: new insights from global analyses. *Cell* **126**: 37-47
- 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. *Theor Appl Genet* **111**: 44-49
- Cavrak VV, Lettner N, Jamge S, Kosarewicz A, Bayer LM, Mittelsten Scheid O** (2014) How a retrotransposon exploits the plant's heat stress response for its activation. *PLoS Genet* **10**: e1004115
- Concibido VC, Diers BW, Arelli PR** (2004) A Decade of QTL Mapping for Cyst Nematode Resistance in Soybean. *Crop Sci.* **44**: 1121-1131
- Cook DE, Bayless AM, Wang K, Guo X, Song Q, Jiang J, Bent AF** (2014) Distinct Copy Number, Coding Sequence, and Locus Methylation Patterns Underlie Rhg1-Mediated Soybean Resistance to Soybean Cyst Nematode. *Plant Physiol* **165**: 630-647
- 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
- Ding M, Ye W, Lin L, He S, Du X, Chen A, Cao Y, Qin Y, Yang F, Jiang Y, Zhang H, Wang X, Paterson AH, Rong J** (2015) The Hairless Stem Phenotype of Cotton (*Gossypium barbadense*) Is Linked to a Copia-Like Retrotransposon Insertion in a Homeodomain-Leucine Zipper Gene (HD1). *Genetics* **201**: 143-154
- Du J, Grant D, Tian Z, Nelson RT, Zhu L, Shoemaker RC, Ma J** (2010) SoyTEdb: a comprehensive database of transposable elements in the soybean genome. *BMC Genomics* **11**: 113
- Du J, Tian Z, Hans CS, Laten HM, Cannon SB, Jackson SA, Shoemaker RC, Ma J** (2010) Evolutionary conservation, diversity and specificity of LTR-retrotransposons in flowering plants: insights from genome-wide analysis and multi-specific comparison. *Plant J* **63**: 584-598
- Galindo-Gonzalez L, Mhiri C, Deyholos MK, Grandbastien MA** (2017) LTR-retrotransposons in plants: Engines of evolution. *Gene* **626**: 14-25

- Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS** (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res* **40**: D1178-1186
- Havecker ER, Gao X, Voytas DF** (2004) The diversity of LTR retrotransposons. *Genome Biol* **5**: 225
- Hewezi T, Lane T, Piya S, Rambani A, Rice JH, Staton M** (2017) Cyst Nematode Parasitism Induces Dynamic Changes in the Root Epigenome. *Plant Physiol* **174**: 405-420
- Kanazawa A, Liu B, Kong F, Arase S, Abe J** (2009) Adaptive evolution involving gene duplication and insertion of a novel Ty1/copia-like retrotransposon in soybean. *J Mol Evol* **69**: 164-175
- Kidwell MG, Lisch DR** (2001) Perspective: transposable elements, parasitic DNA, and genome evolution. *Evolution* **55**: 1-24
- Kim MY, Zilberman D** (2014) DNA methylation as a system of plant genomic immunity. *Trends Plant Sci* **19**: 320-326
- Klock HE, Lesley SA** (2009) The Polymerase Incomplete Primer Extension (PIPE) method applied to high-throughput cloning and site-directed mutagenesis. *Methods Mol Biol* **498**: 91-103
- Krom N, Recla J, Ramakrishna W** (2008) Analysis of genes associated with retrotransposons in the rice genome. *Genetica* **134**: 297-310
- Lee TG, Kumar I, Diers BW, Hudson ME** (2015) Evolution and selection of Rhg1, a copy-number variant nematode-resistance locus. *Mol Ecol* **24**: 1774-1791
- Lisch D** (2009) Epigenetic regulation of transposable elements in plants. *Annu Rev Plant Biol* **60**: 43-66
- Lisch D, Bennetzen JL** (2011) Transposable element origins of epigenetic gene regulation. *Curr Opin Plant Biol* **14**: 156-161
- Lisch D, Slotkin RK** (2011) Strategies for silencing and escape: the ancient struggle between transposable elements and their hosts. *Int Rev Cell Mol Biol* **292**: 119-152
- Liu J, He Y, Amasino R, Chen X** (2004) siRNAs targeting an intronic transposon in the regulation of natural flowering behavior in Arabidopsis. *Genes Dev* **18**: 2873-2878
- Liu S, Kandoth PK, Lakhssassi N, Kang J, Colantonio V, Heinz R, Yeckel G, Zhou Z, Bekal S, Dapprich J, Rotter B, Cianzio S, Mitchum MG, Meksem K** (2017) The soybean GmSNAP18 gene underlies two types of resistance to soybean cyst nematode. *Nat Commun* **8**: 14822

- 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
- Matsunaga W, Kobayashi A, Kato A, Ito H** (2012) The effects of heat induction and the siRNA biogenesis pathway on the transgenerational transposition of ONSEN, a copia-like retrotransposon in *Arabidopsis thaliana*. *Plant Cell Physiol* **53**: 824-833
- Matsunaga W, Ohama N, Tanabe N, Masuta Y, Masuda S, Mitani N, Yamaguchi-Shinozaki K, Ma JF, Kato A, Ito H** (2015) A small RNA mediated regulation of a stress-activated retrotransposon and the tissue specific transposition during the reproductive period in *Arabidopsis*. *Front Plant Sci* **6**: 48
- McCue AD, Nuthikattu S, Reeder SH, Slotkin RK** (2012) Gene expression and stress response mediated by the epigenetic regulation of a transposable element small RNA. *PLoS Genet* **8**: e1002474
- McCue AD, Nuthikattu S, Slotkin RK** (2013) Genome-wide identification of genes regulated in trans by transposable element small interfering RNAs. *RNA Biol* **10**: 1379-1395
- McCue AD, Slotkin RK** (2012) Transposable element small RNAs as regulators of gene expression. *Trends Genet* **28**: 616-623
- Meksem K, Pantazopoulos P, Njiti V, Hyten L, Arelli P, Lightfoot D** (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
- Meyers BC, Tingey SV, Morgante M** (2001) Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. *Genome Res* **11**: 1660-1676
- Mita P, Boeke JD** (2016) How retrotransposons shape genome regulation. *Curr Opin Genet Dev* **37**: 90-100
- Mitchum MG** (2016) Soybean Resistance to the Soybean Cyst Nematode *Heterodera glycines*: An Update. *Phytopathology* **106**: 1444-1450
- Niblack T, Colgrove A, Colgrove K, Bond J** (2008) Shift in virulence of soybean cyst nematode is associated with use of resistance from PI 88788. *Plant Health Prog* **10**
- Niblack TL, Arelli PR, Noel GR, Opperman CH, Orf JH, Schmitt DP, Shannon JG, Tylka GL** (2002) A Revised Classification Scheme for Genetically Diverse Populations of *Heterodera glycines*. *J Nematol* **34**: 279-288

- Niblack TL, Lambert KN, Tylka GL** (2006) A model plant pathogen from the kingdom Animalia: *Heterodera glycines*, the soybean cyst nematode. *Annu Rev Phytopathol* **44**: 283-303
- Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, 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. *Nature* **463**: 178-183
- Schulman AH** (2013) Retrotransposon replication in plants. *Curr Opin Virol* **3**: 604-614
- Slotkin RK, Martienssen R** (2007) Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* **8**: 272-285
- Song J, Keppler BD, Wise RR, Bent AF** (2015) PARP2 Is the Predominant Poly(ADP-Ribose) Polymerase in Arabidopsis DNA Damage and Immune Responses. *PLoS Genet* **11**: e1005200
- Tsuchiya T, Eulgem T** (2013) An alternative polyadenylation mechanism coopted to the Arabidopsis RPP7 gene through intronic retrotransposon domestication. *Proc Natl Acad Sci U S A* **110**: E3535-3543
- Woodrow P, Pontecorvo G, Fantaccione S, Fuggi A, Kafantaris I, Parisi D, Carillo P** (2010) Polymorphism of a new Ty1-copia retrotransposon in durum wheat under salt and light stresses. *Theor Appl Genet* **121**: 311-322
- Xiao H, Jiang N, Schaffner E, Stockinger EJ, van der Knaap E** (2008) A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science* **319**: 1527-1530
- Zhao M, Ma J** (2013) Co-evolution of plant LTR-retrotransposons and their host genomes. *Protein Cell* **4**: 493-501
- Zhao M, Wu S, Zhou Q, Vivona S, Cipriano DJ, Cheng Y, Brunger AT** (2015) Mechanistic insights into the recycling machine of the SNARE complex. *Nature* **518**: 61-67

## Chapter 7: Future Directions

### 7.1 Future Directions

Over many years, the combined efforts of several Bent lab members as represented in Chapters II-VI, has contributed greatly to our understanding of the genetic architecture and molecular function of *Rhg1*. Together, these works build a paradigm by which to view not only *Rhg1*, but one which also unifies and explains how other QTL implicated in SCN resistance operate at a molecular level (Webb et al., 1995; Kopisch-Obuch and Diers, 2006; Matsye et al., 2012; Jiao et al., 2015; Vuong et al., 2015; Lakhssassi et al., 2017). Our studies of *Rhg1* single copy (SCN-susceptible), *Rhg1* high copy, and *Rhg1* low copy soybeans indicate that *Rhg1*-mediated SCN resistance essentially “replaces and/or rewires” the core components of the SNARE-recycling machinery ( $\alpha$ -SNAP and NSF). Multi-copy *Rhg1* haplotypes lose and/or outcompete WT  $\alpha$ -SNAPs by encoding multiple repeat copies of atypical  $\alpha$ -SNAPs, which are impaired in normal NSF interactions. Loss of WT activity of an essential housekeeping protein like  $\alpha$ -SNAP, however, necessitates co-inheritance of an unusual and novel NSF allele (*RAN07*), whose product restores some compatibility with the polymorphic *Rhg1* resistance-type  $\alpha$ -SNAPs. Nonetheless, these discoveries aside, many basic and intriguing questions regarding *Rhg1* function remain unanswered.

Perhaps foremost of these questions is, what are the other *Rhg1*-encoded genes doing? Conversely, if just the *Rhg1*  $\alpha$ -SNAPs are sufficient for resistance, might combining both allelic variants ( $\alpha$ -SNAP<sub>*Rhg1*HC</sub> and  $\alpha$ -SNAP<sub>*Rhg1*LC</sub>) bolster resistance, while increasing WT  $\alpha$ -SNAP abundance would deplete resistance? Cook *et al* 2012 indicated that silencing either of two additional *Rhg1* genes (*Glyma18g02580* - amino acid permease; *Glyma18g02610* - wound-inducible protein) increased SCN susceptibility, while co-expression of these two genes with  $\alpha$ -

SNAP<sub>Rhg1HC</sub> was also needed for enhanced SCN-resistance in transgenic roots (Cook et al., 2012). Moreover, Cook *et al* 2014 demonstrated that *Rhg1* copy number gains increased mRNA abundance of these genes, which also correlated with improved SCN resistance (Cook et al., 2014). However, a recent study suggests that  $\alpha$ -SNAP<sub>Rhg1LC</sub> alone can confer SCN-resistance (Liu et al., 2017). Additionally, transgenic soybean plants expressing epitope-tagged versions of these three genes, including  $\alpha$ -SNAP<sub>Rhg1HC</sub>, did not elevate SCN resistance compared to SCN-susceptible varieties (not shown). These plants, however, were generated prior to our discovery of *RAN07* and immunoblots suggested that expression of the *Rhg1*-encoded proteins was weak while epitope tags could impact function (not shown).

The inherently low sensitivity of SCN resistance assays using transgenic hairy roots has complicated answering if these additional genes assist resistance or if they are merely linked within the *Rhg1* block containing the resistance-type  $\alpha$ -SNAPs. To resolve this central question, we are generating stable transgenic soybeans (via the Wisconsin Crop Innovation Center) of *Rhg1<sub>HC</sub>*, *Rhg1<sub>LC</sub>* and SCN-susceptible backgrounds, in which each of the other *Rhg1*-encoded genes are silenced via RNAi. To this end, I generated shRNA constructs targeting each of these three additional *Rhg1* repeat genes. Additionally, I assembled constructs expressing untagged  $\alpha$ -SNAP<sub>Rhg1HC</sub>,  $\alpha$ -SNAP<sub>Rhg1LC</sub>, or  $\alpha$ -SNAP<sub>Rhg1WT</sub> that also co-express NSF<sub>Ch07</sub> or RAN07. These will be transformed into the same *Rhg1<sub>HC</sub>*, *Rhg1<sub>LC</sub>* or SCN-susceptible backgrounds. Immunoblots will assess resistance-type and WT  $\alpha$ -SNAP expression in all lines, and in the silencing experiments, qPCR can verify knockdown of the other *Rhg1*-encoded genes. Finally, SCN cup assays can determine the SCN resistance phenotype of these transgenic soybean plants. Using whole plants, these assays are anticipated to answer if 1) silencing *Rhg1* genes outside of  $\alpha$ -SNAP impacts resistance, 2) if *Rhg1* resistance-type  $\alpha$ -SNAP expression alone can confer

detectable SCN resistance, 3) if stacking both  $\alpha$ -SNAP<sub>Rhg1HC</sub> and  $\alpha$ -SNAP<sub>Rhg1LC</sub> alleles enhances overall resistance and if to particular SCN populations, and 4) if elevated WT  $\alpha$ -SNAP expression in already SCN-resistant backgrounds reduces *Rhg1* efficacy. Moreover, these transgenic plants will demonstrate if RAN07 vs. WT NSF<sub>Ch07</sub> is needed for the viability of *Rhg1* resistance-type  $\alpha$ -SNAP expressing plants.

The observed decrease of WT  $\alpha$ -SNAP abundance observed in *Rhg1LC* varieties (Ch V) was striking. Moreover, immunogold-EM (electron microscopy) labeling indicated that  $\alpha$ -SNAP<sub>Rhg1HC</sub> hyperaccumulates precisely within nascent syncytial cells as compared with adjacent cells, while immunoblots suggested that WT  $\alpha$ -SNAPs do not increase as much as  $\alpha$ -SNAP<sub>Rhg1HC</sub>. Put together, these findings support that WT  $\alpha$ -SNAPs are important factors in SCN syncytium formation and that depleting/out-competing WT  $\alpha$ -SNAP activity is a key element of *Rhg1*-mediated SCN resistance.

The question then arises, why do WT  $\alpha$ -SNAPs benefit SCN? But really, this is a question of syncytium biology examined in terms of NSF and  $\alpha$ -SNAP dynamics in the SCN-compatible (susceptible) and in the SCN-resistant backgrounds (*Rhg1HC* and *Rhg1LC*). Understanding the role of WT  $\alpha$ -SNAPs (and NSF) during syncytium formation in the compatible interaction will likely facilitate a deeper understanding of how *Rhg1* resistance perturbs this function. Unlike EM labeling, immunoblots did not precisely reveal  $\alpha$ -SNAP<sub>Rhg1HC</sub> location or abundance in syncytium cells. Unfortunately, due to non-specificity of the WT  $\alpha$ -SNAP antibody with EM labeling, an accurate quantification and localization of WT  $\alpha$ -SNAPs in syncytia is currently untenable. However, in Ch V, transgenic addition of the native WT *Rhg1*  $\alpha$ -SNAP locus boosted WT  $\alpha$ -SNAP abundance similar to levels observed in SCN-susceptible backgrounds. Therefore, an immunogold-compatible epitope tag could be inserted within the

native  $\alpha$ -SNAP ORF (via Gibson assembly), thereby bypassing the WT- $\alpha$ -SNAP antibody, while maintaining any *cis*-regulatory elements within the native locus. This native tagged  $\alpha$ -SNAP could be transformed into SCN-susceptible and resistant soybeans to examine WT *Rhg1*  $\alpha$ -SNAP induction and localization within syncytial as compared to neighboring cells, and importantly, if similar dynamics occur within syncytia of SCN-resistant soybeans. Additionally, *Arabidopsis* and *Heterodera schachtii* models could be useful in examining the necessity of WT  $\alpha$ -SNAPs and NSF in syncytia. This system would benefit from the use of existing lines with fluorescently labeled cellular compartments that could monitor syncytium progression, as well as the ease of creating new transgenic lines (or CRISPR) and amenability to live cell confocal imaging.

The question of if the abundance of RAN07 vs. WT NSF<sub>Ch07</sub> vs. NSF<sub>Ch13</sub> fluctuates during SCN-syncytium formation is also of interest. If high levels RAN07 are present within the syncytium, this would support that resistance-type  $\alpha$ -SNAPs are actually pro-vesicle trafficking, while if RAN07 decreases, this would suggest that as proposed in Bayless *et al*, resistance-type  $\alpha$ -SNAPs function via a dominant negative function. A RAN07 specific antibody will be essential for future studies, and ideally, this antibody should be compatible by both immunoblot and EM-labeling studies. Because immunoblots poorly revealed the precise location and abundance of  $\alpha$ -SNAP<sub>*Rhg1*HC</sub> within syncytium cells, investigations of NSF and RAN07 should utilize EM or at least laser capture microdissection for determinations of protein abundance.

The necessity of *RAN07* as an apparent balance for multi-copy *Rhg1* haplotypes and the lack of WT  $\alpha$ -SNAPs in *Rhg1* low copy backgrounds questions the semi-dominant negative model proposed in Bayless *et al*. This model could still be valid for *Rhg1*<sub>HC</sub> varieties, and is somewhat supported by loss of functional Ch11 WT  $\alpha$ -SNAP loci as boosting resistance, but appears untenable for *Rhg1*<sub>LC</sub> varieties.  $\alpha$ -SNAP<sub>*Rhg1*LC</sub> likely is performing some housekeeping

maintenance of SNARE disassembly given the dearth of WT  $\alpha$ -SNAP abundance. While RAN07 was more effective at binding resistance type  $\alpha$ -SNAPs and rescuing cell death, further questions regarding RAN07 remain. For instance, does RAN07, in conjunction with *Rhg1* resistance-type  $\alpha$ -SNAPs assist resistance, or is RAN07 a passive player that simply promotes the viability of *Rhg1* germplasms? Additionally, with RAN07, do *Rhg1* resistance-type  $\alpha$ -SNAPs function as effectively as WT  $\alpha$ -SNAPs? Does RAN07 work equally well with  $\alpha$ -SNAP<sub>*Rhg1*HC</sub> vs.  $\alpha$ -SNAP<sub>*Rhg1*LC</sub>, or do subtle functional differences exist? Biochemical assays investigating how the different *Rhg1*  $\alpha$ -SNAPs affect ATPase stimulation, and SNARE-bundle disassembly of RAN07 vs. NSF<sub>Ch07</sub>, could be explored. Additionally, the native WT  $\alpha$ -SNAP locus of *Rhg1*<sub>HC</sub> varieties could be selectively targeted using CRISPR, and if these transgenic plants are non-viable, it would suggest that WT  $\alpha$ -SNAPs are needed for viability in *Rhg1*<sub>HC</sub> backgrounds.

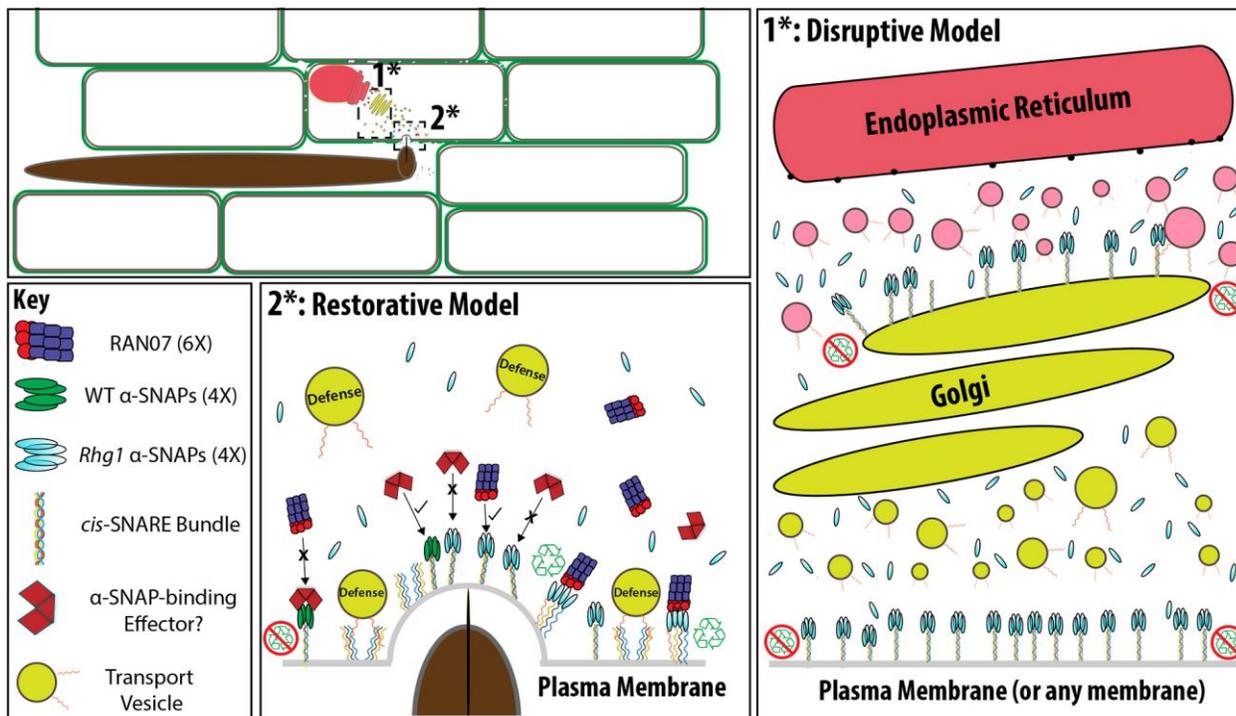
If  $\alpha$ -SNAP<sub>*Rhg1*HC</sub> vs.  $\alpha$ -SNAP<sub>*Rhg1*LC</sub> essentially just don't interact well with normal NSF, how are they different? Curiously, these high vs. low copy *Rhg1* sources are more effective against particular SCN populations (Brucker et al., 2005). Why was there an apparent allelic divergence after the original *Rhg1* duplication event – could SCN selective pressures be maintaining these two different  $\alpha$ -SNAP alleles? Having both *Rhg1* alleles could be beneficial to soybean perhaps if SCN maintains an effector that normally targets WT  $\alpha$ -SNAPs. Having *Rhg1* resistance-type  $\alpha$ -SNAPs would diminish this effector interaction/targeting and thus reduce potential manipulation by SCN. This would also explain why getting rid of WT  $\alpha$ -SNAPs is also advantageous. Targeting a conserved interface between two housekeeping proteins underlying a central cellular pathway (vesicle trafficking) could be an effective virulence strategy for a pathogen. Moreover, this effector should work broadly among different hosts given the high conservation of  $\alpha$ -SNAP and NSF interactions. When SCN populations are reared on *Rhg1* HC

hosts, the effector could be selected to better target  $\alpha$ -SNAP<sub>RhgI</sub>HC, but not  $\alpha$ -SNAP<sub>RhgI</sub>LC, and vice versa (Gardner et al., 2017). Effector sequence mining and examining allelic differences among virulent SCN populations raised on particular sources of *RhgI* resistance could provide further support for this speculation. Yeast two hybrid assays or co-IP could then be used to explore putative effector -  $\alpha$ -SNAP or NSF interactions. If such an effector exists and is key to the pathogen-host interaction, then stacking  $\alpha$ -SNAP alleles might be an effective defense strategy. Moreover, the possibility of synthetic  $\alpha$ -SNAP and NSF alleles which still interact and disassemble SNAREs, but evade effector targeting could also be explored.

Lastly, is there a biological significance for why *RhgI*<sub>LC</sub>, but not *RhgI*<sub>HC</sub> haplotypes harbor an intact Copia retrotransposon (*RAC*)? Could lack of WT  $\alpha$ -SNAP expression in *RhgI*<sub>LC</sub> backgrounds require stringent control of  $\alpha$ -SNAP<sub>RhgI</sub>LC under certain conditions? *RAC* is anti-sense within intron 1 of  $\alpha$ -SNAP<sub>RhgI</sub>LC – could *RAC* act as a domesticated transposon and provide a host benefit by regulating  $\alpha$ -SNAP<sub>RhgI</sub>LC? If so, when and how might *RAC* act? These questions are not only intriguing, but also of interest to the agricultural community should *RhgI*<sub>LC</sub> haplotypes become more widely used. In Ch VI, we detected no impacts of *RAC* upon  $\alpha$ -SNAP<sub>RhgI</sub>LC, however, further experimentation should be performed. For example, assessing if *RAC* can initiate anti-sense transcription against the  $\alpha$ -SNAP<sub>RhgI</sub>LC ORF would be a step towards determining if *RAC* might regulate  $\alpha$ -SNAP<sub>RhgI</sub>LC expression. To this end, I cloned only the native *RAC* element into a binary expression construct. Transforming this *RAC* only construct into Wm82 (low levels of endogenous *RAC*-like transcripts) and then assessing *RAC*-like transcript levels will answer if *RAC* can initiate transcription and to what extent. Furthermore, exploring if the  $\alpha$ -SNAP<sub>RhgI</sub>LC transcript splices within regions of *RAC* is of potential interest. Northern blotting and/or cDNA analysis could examine if such transcripts exist. Also, examining

if other stresses, like drought, cold, salt, or in particular, pathogen (i.e., SCN) impact *RAC* transcription or epigenetic modifications, such as methylation or histone modifications at and near *RAC*, and in turn, if these impact  $\alpha$ -SNAP<sub>*RhgI*LC</sub> expression, is of interest. Much about *RAC* remains unanswered. Regardless, future examination of how SCN infection might modulate resistance-type  $\alpha$ -SNAP expression, in both *RhgI*<sub>HC</sub> and *RhgI*<sub>LC</sub> varieties, is of considerable interest.

## 7.2 Model



### Models of *Rhg1* $\alpha$ -SNAP mediated SCN-resistance.

Depending on *Rhg1* resistance-type  $\alpha$ -SNAP and RAN07 protein abundance in the SCN-induced syncytium, two models are possible: the “Disruptive” and “Restorative”.

In 1\* (Disruptive Model), *Rhg1* resistance-type  $\alpha$ -SNAPs hyperaccumulate within syncytial cells and actively disrupt trafficking and membrane fusion events. A key tenet of the model is that RAN07 protein abundance would not increase within the syncytium, which is currently unknown, and while RAN07 is enhanced for resistance-type  $\alpha$ -SNAP binding, it is unclear if RAN07/resistance-type  $\alpha$ -SNAP mediated SNARE-disassembly is as efficient as WT- $\alpha$ -SNAP/WT-NSF pairs. SNARE complexes on any target membrane (only showing Golgi and plasma membrane) are not efficiently disassembled and vesicular trafficking and/or membrane fusion events are severely hampered. This compromises syncytium development and leads to

collapse, especially as the nematode exerts greater physiological and metabolic demands and as new neighboring cells are incorporated into the syncytium.

In 2\* (Restorative Model), *Rhg1* resistance-type  $\alpha$ -SNAPs hyperaccumulate and promote restoration of vesicle trafficking. In this model, SCN utilizes an effector to actively undermine host cell trafficking networks, and it does this by binding to the conserved C-terminus of WT  $\alpha$ -SNAPs. This effector would be essential for cyst nematode parasitism and could benefit SCN by blocking immune trafficking (shown) but also by blocking  $\alpha$ -SNAP and NSF function anywhere in the syncytium. Because  $\alpha$ -SNAP and NSF are core housekeeping chaperones, binding  $\alpha$ -SNAPs would effectively shut down trafficking anywhere in the cell. Directly binding  $\alpha$ -SNAPs could also promote non-specific fusion events required for syncytia formation via reducing SNARE-proofreading. This model could also explain certain SCN-virulence populations (i.e., their allele of this effector can somewhat bind certain resistance-type  $\alpha$ -SNAPs as well as WT  $\alpha$ -SNAPs). Perhaps more importantly, this model also accounts for why *Rhg1* resistance selects against WT  $\alpha$ -SNAPs – they are easily targeted and inactivated by the effector, and why it is beneficial to *Rhg1* resistance to have different resistance type  $\alpha$ -SNAP alleles. I would speculate that this model reflects how resistance-type  $\alpha$ -SNAPs function during resistance. Post-translational modifications, especially those induced by SCN infection are unknown and could impact  $\alpha$ -SNAPs and NSF (RAN07) interactions.

### 7.3 References

- 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. *Theor Appl Genet* **111**: 44-49
- Cook DE, Bayless AM, Wang K, Guo X, Song Q, Jiang J, Bent AF** (2014) Distinct Copy Number, Coding Sequence, and Locus Methylation Patterns Underlie Rhg1-Mediated Soybean Resistance to Soybean Cyst Nematode. *Plant Physiol* **165**: 630-647
- 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
- Gardner M, Heinz R, Wang J, Mitchum MG** (2017) Genetics and Adaptation of Soybean Cyst Nematode to Broad Spectrum Soybean Resistance. *G3 (Bethesda)* **7**: 835-841
- Jiao Y, Vuong TD, Liu Y, Meinhardt C, Liu Y, Joshi T, Cregan PB, Xu D, Shannon JG, Nguyen HT** (2015) Identification and evaluation of quantitative trait loci underlying resistance to multiple HG types of soybean cyst nematode in soybean PI 437655. *Theor Appl Genet* **128**: 15-23
- Kopisch-Obuch FJ, Diers BW** (2006) Segregation at the SCN resistance locus rhg1 in soybean is distorted by an association between the resistance allele and reduced field emergence. *Theor Appl Genet* **112**: 199-207
- Lakhssassi N, Liu S, Bekal S, Zhou Z, Colantonio V, Lambert K, Barakat A, Meksem K** (2017) Characterization of the Soluble NSF Attachment Protein gene family identifies two members involved in additive resistance to a plant pathogen. *Sci Rep* **7**: 45226
- Liu S, Kandoth PK, Lakhssassi N, Kang J, Colantonio V, Heinz R, Yeckel G, Zhou Z, Bekal S, Dapprich J, Rotter B, Cianzio S, Mitchum MG, Meksem K** (2017) The soybean GmSNAP18 gene underlies two types of resistance to soybean cyst nematode. *Nat Commun* **8**: 14822
- 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
- Vuong TD, Sonah H, Meinhardt CG, Deshmukh R, Kadam S, Nelson RL, Shannon JG, Nguyen HT** (2015) Genetic architecture of cyst nematode resistance revealed by genome-wide association study in soybean. *BMC Genomics* **16**: 593

**Webb DM, Baltazar BM, Rao-Arelli AP, Schupp J, Clayton K, Keim P, Beavis WD (1995)**  
Genetic mapping of soybean cyst nematode race-3 resistance loci in the soybean PI  
437.654. *Theor Appl Genet* **91**: 574-581