Investigating *Rhg1* in the annual *Glycine* species *Glycine max* and *Glycine soja*

By

Derrick J. Grunwald

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Microbiology)

at the

UNIVERSITY OF WISCONSIN-MADISON

2021

Date of final oral examination: 01/05/2021

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

Andrew F. Bent, Professor, Plant Pathology
Caitilyn Allen, Professor, Plant Pathology
Sebastian Bednarek, Professor, Biochemistry
J. D. Sauer, Associate Professor, Medical Microbiology and Immunology
Jean-Michel Ané, Professor, Plant Pathology

Acknowledgments

I would first like to thank my parents, Edward Grunwald and Sue Levin, for their support along the way. Even though you may not have understood what I am doing or why I am doing it, especially in Wisconsin, you have provided nothing but encouragement along the way. For this, I am incredibly grateful.

I would also like to thank my advisor, Dr. Andrew Bent, for providing me a lab home, and giving me the opportunity to develop as a scientist, mentor and educator in my own right. Whether providing a forum to discuss scientific ideas, sharing memories about our respective times in Berkeley and in the Staskawicz lab, or hearing about his love for soup and lemon cake. Andrew has time and again proven himself to be an ideal mentor. For his encouragement, patience, and open-mindedness, I will forever be grateful to him. Additionally, I would like to thank my current thesis committee members: Drs. Caitilyn Allen, Jean-Michel Ané, Sebastian Bednarek, and JD Sauer, and past advisor, Dr. Amy Charkowski, for their mentorship along the way. I would also like to thank our Dr. Brian Diers for his collaboration and insights into *Glycine soja* and resistance.

To all the Bent lab members, past and present: Katelyn Butler, Ryan Zapotocny, Hu Die, Yulin Du, Brian Keppler, Shaojie Han, Aaron Lowenstein, Deepak Haarith, Adam Bayless, and Jeysika Zayas-Rivera—you have all made the lab a great place to work, and I know you will all have success in your own rights. I would especially like to thank Adam Bayless, Shaojie Han, and Ryan Zapotocny for making my transition to the lab relatively painless, and each has contributed substantially to our work. I would also like to thank undergraduate students past

and present in the lab: Jarett, Huarui, Austin, Anders, Jordyn, Demi, Jessica, Emma, Kaela, Molly, and Abbie for cleaning up our messes as well as contributing to our research.

A big thank you to the Microbiology Doctoral Training Program and the Department of Plant Pathology for providing a truly supportive community. I would be fortunate to work in a community as bright and supportive a community as this again. A heartfelt thank you to my incredible cohort: Lindsey Bohr, Laura Alexander, Connor Hendrich, Julian Bustamante, Bayleigh Benner, Madison Cox, Mengyao Niu, Jared Erickson, Courtney McDougal, Krizia Perez-Medina, and Kelsey Barrett. I have enjoyed struggling and succeeding with each of you over the past years, and I can't wait to see what is next.

Investigating *Rhg1* in the annual *Glycine* species *Glycine max* and *Glycine soja*

Derrick J. Grunwald

Under the supervision of Professor Andrew F. Bent

At the University of Wisconsin-Madison

Abstract

Soybean (*Glycine max*) is a globally important oilseed and protein crop. The world's most damaging pathogen of soybean is soybean cyst nematode (*Heterodera glycines*), routinely causing over $1 billion USD in US yield losses each year. Management of this pathogen relies on the commonly deployed *Resistance to Heterodera glycines 1* (*Rhg1*) resistance QTL. *Rhg1* is a multi-gene locus that displays copy number variation, with $\leq$3 copies being termed *rhg1-a* and $\geq$4 being termed *rhg1-b*. Encoded in one of the genes of the locus is an α-SNAP (alpha Soluble NSF Attachment Protein) which has an unusual set of C-terminal amino acids that disrupt its ability to interact with NSF (N-ethylmaleimide Sensitive Factor), leading to rapid degeneration and collapse of nematode feeding sites. This thesis aims to further characterize both the *Rhg1* locus in annual soybean (*G. max* and *G. soja*) and the molecular interactions of the *Rhg1*-encoded α-SNAPs.

We first sought to determine whether there were any other loci in *G. max*, which co-associate with *Rhg1* in resistant varieties. We found that all resistant-type *G. max* had a chromosome 7 QTL that contained a polymorphic NSF, termed RAN07 (*Rhg1*-Associated NSF on Chromosome 7). Experiments *in vitro* demonstrated that RAN07 is able to interact with

resistance-type α-SNAPs better than a wild-type NSF. Subsequent *in planta* studies demonstrated that RAN07 is required for the presence of *Rhg1*, an observation utilized in later studies. Additionally, soybean varieties with *rhg1-a* often carry a QTL on chromosome 11, which we found is associated with an intron retention allele that reduces the overall abundance of "wild-type" α-SNAP.

We next sought to characterize *rhg1-a* in *G. max*, focusing on the discovery that *rhg1-a* haplotypes contain a retrotransposon in the first intron of the *Rhg1* α-SNAP. With chapter 2, this suggests that there is coordinated regulation of α-SNAP and/or NSF in resistance reactions to soybean cyst nematode in *G. max*.

We finally characterize *Rhg1* in *G. soja*, the wild, annual relative of *G. max*. We first discovered a unique haplotype of *Rhg1* in *G. soja*, that seems to predate the split of *rhg1-a* and *rhg1-b* in *G. max*. We termed this *Rhg1* haplotype *rhg1-cs*. Further characterization of *rhg1-cs* revealed that the encoded α-SNAP gene is distinct. We additionally performed copy number and resistance assays on select *G. soja*. Interestingly, the selected *G. soja* are resistant to highly virulent nematode populations without utilizing known resistance mechanisms to confer this resistance. Finally, we determined the presence of a novel α-SNAP on chromosome 11 of one of the resistant *G. soja* varieties.

The results of this thesis further expand upon the paradigm of *Rhg1*-mediated resistance and provide novel *G. soja* germplasm of both academic and practical interest. Such work provides the foundation for further studies into the intricacies of *Rhg1*-mediated resistance and its evolution.

Chapter 1: Introduction

1.1 A brief Introduction to soybean

Soybean (*Glycine max*) is a global crop, rich in both protein and oil. Soybean production is an important supplier of soybean meal for animal feed and for human consumption (Hartman et al., 2011; FAO stats). Most soybean (nearly 81%) is produced in three principal growing regions: Brazil, the United States, and Argentina (soystats.com). In the US, nearly 3.56 billion bushels of soybean were produced in 2019, with a total value of approximately 31.2 billion USD (soystats.com). The nutrient density of soybean protein and oil, and its content of essential vitamins and minerals, also make soybean an attractive crop in subsistence communities that may otherwise be food insecure (Hartman et al., 2011).

Soybean domestication is known to derive from *Glycine soja*. The route of domestication of soybean remains an open question in the literature, with three competing hypotheses. Domestication could have arisen through a single domestication event from *G. soja* in China 9000 years ago (Zhou et al., 2015b; Kofsky et al., 2018). Such a hypothesis is supported by resequencing of 300+ *G. soja* and comparing the nucleotide diversity at multiple domestication loci to domesticated soybean (Zhou et al., 2015b). Alternatively, soybean could have been domesticated independently at many different locales, with the timeframe for domestication being between 5000-9000 years ago (Kofsky et al., 2018). Such a hypothesis is supported by assembly of a pan-genome for *G. soja* from seven geographically distinct accessions (Li et al., 2014). Interestingly, while most of the genome was shared between accessions, nearly one fifth of the genome experienced regional variation in genome

architecture either relating to copy number variation of genes or their presence/absence (Li et al., 2014). A third hypothesis asserts that the *G. max* domestication from *G. soja* occurred only once, but there was further regional selection from a derived *G .max*/*G. soja* complex during the domestication process that led to unique subpopulations (Kofsky et al., 2018). This model takes a somewhat middle ground approach to the other two and is supported by a synthesis of the data. Further, this latter hypothesis of domestication suggests that the age of this *G. max*/*G. soja* complex to be between 0.27 and 0.8 million years ago, depending on the number of accessions included in the analysis. Whole genome comparison of one accession leads to an estimate of 0.27 mya (Kim et al., 2010b) and comparison of single nucleotide polymorphisms of 302 accessions leads to an estimate of 0.8 mya (Li et al., 2014). Importantly, the sexual compatibility of *G. soja* with primarily self-fertilizing *G. max*, combined with whole genome sequence data available for both, enables gene discovery and breeding for biotic and abiotic stresses (Kim et al., 2010b; Schmutz et al., 2010).

1.2 SCN: Biology, Effectors, and Control Strategies

Soybean production is limited by both abiotic and biotic stresses. Nematodes, particularly cyst and root knot nematodes, represent a major biotic threat to plant health. Soybean cyst nematode (*Heterodera glycines*; SCN) is a global pathogen, broadly distributed across the globe where soybeans are produced (EPPO, 2014) and is spreading (Wang et al., 2015; Peng et al., 2016). In the US, SCN was introduced through North Carolina before 1955 (Winstead et al., 1955) and has since spread to all soybean producing states in the US and into parts of Canada (Tylka and Marett, 2017). In terms of production, SCN routinely causes

hundreds of millions of lost bushels at the national scale and may locally contribute to yield losses upwards of 30% (Chen et al., 2001b; Koenning and Wrather, 2010; Jones et al., 2013; Allen et al., 2017). The total farmgate economic cost of SCN is at least 1 billion USD (Koenning and Wrather, 2010; Allen et al., 2017), but the actual economic cost may be much higher when accounting for the costs of diagnostics, treatment, rejected seed and when compounded with abiotic factors (Allen et al., 2017). This makes SCN an attractive target for study.

SCN is an obligate endoparasite of a subset of legume species including both soybeans and common bean. The lifecycle of SCN has been characterized under gnotobiotic conditions in Lauritis et al., 1983 (Figure 1). The nematode undergoes a first molt within the egg, emerging as a J2 (Lauritis et al., 1983; reviewed in Niblack et al., 2006; Turner and Rowe, 2006). A successful J2 then migrates to the root zone of elongation, where it selects a single cell (often a cortical cell adjacent to the vascular bundle) to initiate a nematode feeding structure, termed a syncytium (Niblack et al., 2006; Turner and Rowe, 2006). The nematode then undergoes subsequent molts into a J3. After sexual differentiation and molting into a J4, male nematodes regain their vermiform shape and exit the root to search for females, while females remain at a syncytium (Lauritis et al., 1983; Niblack et al., 2006). Fertilization occurs after about 12 days, and the fertilized female's body cavity fills with eggs (Lauritis et al., 1983). The female body typically extends out from the root and their lemon-shaped bodies become visible on root surfaces. Upon maturity, the female body hardens and melanizes becoming a protective cyst for the eggs inside (Niblack et al., 2006). After about 19 days, the juvenile nematodes become visible within the eggs, meaning that the entire lifecycle of the nematode takes a minimum of about 21 days from J2 to J2 under gnotobiotic conditions (Lauritis et al., 1983). Under field

conditions there is often more than one lifecycle during the season, but hatching signals, soil temperature, humidity, and alkalinity can significantly limit the number of generations (Chen et al., 2001b; Niblack et al., 2006).

The formation of the nematode feeding structure—the syncytium—is a highly orchestrated process, requiring the concerted activities of both nematode and plant host (Kyndt et al., 2013). Broadly speaking, the syncytium is characterized by dense host cytosol, no central vacuole, and enlarged nuclei (Figure 2; de Almeida Engler and Gheysen, 2013; Kyndt et al., 2013). Recent research suggests additional roles of plant cell cycle checkpoints (particularly the cyclin and cyclin dependent kinases CDC2a/b, CYCB1/A2), as well as both a component of the topoisomerase decatenation complex and tubulin in developing syncytia (de Almeida Engler et al., 1999; de Almeida Engler and Gheysen, 2013; Kyndt et al., 2013). Syncytial maintenance seems to require the hijacking of plant-cell hormones or small molecules. Indeed, evidence suggests a role of auxin and cytokinin in promoting syncytial development, and there is even a nematode-produced mimic of the CLAVATA3-recognized CLE peptide, part of the *wuschel* pathway, to promote differentiation (Grunewald et al., 2009; Wang et al., 2010; Wang et al., 2010b; Siddique et al., 2015). Nematodes produce many additional small molecules and proteins to promote virulence (Mejias et al., 2019).

At a broad scale, SCN interactions with the plant are determined by the effector suites the nematode uses to promote virulence. The difficulty in manipulating nematodes has often restricted the study of effectors to stable expression in plant lines. In particular, use of the *Arabidopsis* and beet cyst nematode (*Heterodera schachtii*, BCN) pathosystem has greatly accelerated the functional characterization of many nematode effectors (Hewezi and Baum,

2013; Mitchum et al., 2013). Through this system, many BCN effectors have been characterized to be important for virulence including ones that mimic plant signaling molecules, ones that modulate expression of plant hormones, ones that degrade host cell wall components, SNARE like proteins, and even ones that may modify plant basal defenses (Smant et al., 1998; Bekal et al., 2003; Qin et al., 2004; Wang et al., 2010; Wang et al., 2010b; Gheysen and Mitchum, 2011; Hamamouch et al., 2012; Bekal et al., 2015; Hewezi et al., 2015; Wang et al., 2020). Furthermore, nematode-induced small RNA production has been implicated in producing virulence outcomes by modifying plant growth factor expression (Hewezi et al., 2008). The recent publication of the SCN genome, combined with sequenced effector families as well as DNA sequence motifs associated with secreted effectors, will greatly facilitate discovery of polymorphic or novel effectors from diverse nematode populations (Eves-van den Akker and Birch, 2016; Gardner et al., 2018; Masonbrink et al., 2019).

Cyst longevity in soils as well as the ease of movement of the cysts by contaminated soil, water, wind or machinery has made management of SCN difficult. First generation nematicides such as methyl bromide have largely been discontinued due to their acute toxicity and prohibitive cost of application as a soil fumigant. There have been recent efforts to develop nematicides with both reduced toxicity as well increased efficacy as a seed treatment (Gaspar et al., 2014; Lahm et al., 2017). Additionally, there has been increasing research into the use of bacterial and fungal seed treatments to both prevent nematode growth and improve yields (Zhou et al., 2017; Haarith et al., 2020; Storm et al., 2020). There has been additional suggestion that soil amendments, like liquid swine manure, may impede nematode reproduction (Bao et al., 2013). In many of the above cases, however, field trials have produced

disappointing results suggesting that there may be confounding factors that could limit efficacy of any of the above treatments.

Management of SCN, therefore, has been most successful through the cultural practices of crop rotation and planting resistant varieties. Some of the strongest ways in theory to prevent SCN infestation are to prevent initial introduction through stringent cultural practices: quarantining infested land, banning planting of host crops on uninfected land, suppressing population on infested fields, and preventing shipment of infested materials (Nicol et al., 2011). Consistently, in soybean agrosystems, SCN pressure decreases with rotation to corn; this also has the benefit of reducing pathogen pressure on corn (Grabau and Chen, 2016). There are many commercial cultivars with claimed resistance to SCN. Of these, the *resistance to Heterodera glycines 1* (*Rhg1*) locus is by far the most widely employed and efficacious (Webb et al., 1995; Brucker et al., 2005; Mitchum, 2016; Rincker et al., 2017). RNA interference, through genetic modification or exogenous application, has been an attractive future direction for control strategies but has not yet become an established mode of SCN control (Sindhu et al., 2008; Li et al., 2010).

As noted above, *Rhg1* is a widely deployed locus that confers resistance to SCN, which has led to challenges in management. Resistance in most field-grown soybeans is derived from a single resistance source, the *rhg1-b* haplotype derived from PI 88788 (McCarville et al., 2017). Due to its many favorable traits but (over)use, nearly 97% of soybean varieties grown in either Iowa or Illinois have resistance derived from this plant introduction (Tylka, 2016; McCarville et al., 2017). Use and overuse of this resistance locus is further exemplified by virulence shifts in nematode populations where PI 88788-derived soybeans are grown. Although *rhg1-b* was

successfully used for a few decades, recent field tests in Illinois, Missouri, Kentucky, Tennessee, Ontario, South Dakota, and Indiana demonstrate that an increasing number of field populations of SCN are able to reproduce robustly on soybean varieties that carry the PI 88788-derived *rhg1-b* haplotype (Mitchum et al., 2007; Hershman et al., 2008; Niblack et al., 2008; Faghihi et al., 2010; Acharya et al., 2016; McCarville et al., 2017). Consequently, PI 88788-derived soybeans are yielding less in high-SCN pressure fields (McCarville et al., 2017). There are other sources of resistance, although with less favorable agronomic traits. SCN resistance derived from Peking (PI 548402) has *rhg4* (in addition to the *rhg1-a* haplotype) and is able to confer resistance to the field populations of SCN able to overcome PI 88788-derived resistance (Colgrove and Niblack, 2008). Additionally, PI 437654 has resistance QTLs that are able to confer resistance to nematode populations able to overcome resistance by either PI 88788 or Peking (Mitchum et al., 2007; Hershman et al., 2008; Acharya et al., 2016). However, soybean lines carrying SCN resistance from these sources tended to have lower yields than lines carrying *rhg1-b*, making them less attractive to growers. Continued breeding effort to move these and other alternative resistance QTLs, including those from crop wild relatives (discussed below), into elite-yielding varieties is an attractive option for growers to both produce high-yielding and highly SCN-resistant varieties.

1.3 *Rhg1*: Overview

Rhg1 SCN-resistant soybeans cause rapid degeneration of syncytia between 3-5 days after infection (Mitchum, 2016). Initial attempts to isolate the responsible genes identified an LRR-RLK as most likely to be responsible for SCN resistance (Lightfoot and Meksem, 2001).

Subsequent mapping studies, however, demonstrated that the LRR-RLK was outside of the QTL (Kim et al., 2010), and RNAi functional tests failed to detect a contribution of the *rhg1-b*-linked LRR-RLK to SCN resistance (Melito et al., 2010). Early functional studies on *Rhg1* identified differential expression of stress and defense related genes in *Rhg1*-containing near isogenic lines relative to those lines that did not contain *Rhg1* (Kandoth et al., 2011).

The *Rhg1* locus is complicated, containing four non-canonical defense-related genes and displaying copy number variation (Cook et al., 2012). In particular, within the locus are a putative amino acid permease (*Glyma.18g022400*), an alpha-soluble NSF attachment protein (α-SNAP; *Glyma.18g022500*), a PLAC8 domain containing protein (*Glyma.18g022600*) and a putative wound inducible protein (*Glyma.18g022700*) (Cook et al., 2012). Silencing of any of the three genes: the amino acid permease, the α-SNAP, and the wound inducible protein, reduced resistance of Fayette (PI 88788-derived soybean) to SCN, suggesting a role for these proteins in resistance (Cook et al., 2012). There are further two distinct haplotypes of *Rhg1* that differ in their resistance to different populations of SCN: *rhg1-a* from Peking and *rhg1-b* from PI 88788 (Niblack et al., 2002; Brucker et al., 2005; Mithcum, 2016). These two haplotypes, it was found, also differ in the number of tandemly duplicated copies of *Rhg1* where *rhg1-a* tends to have two or more typically three identical tandemly duplicated copies of the locus and *rhg1-b* has more than four, and often nine or ten tandem duplicate copies of the four-gene locus (Cook et al., 2012). In rhg1-b haplotypes, two-thirds of one of the blocks is similar to susceptible single-copy *Rhg1* soybean varieties such as Williams82 (Cook et al., 2012; Cook et al., 2014; Lee et al., 2015b). Moreover, there are nonsynonymous mutations in the C-terminal region of the α-SNAP genes of resistant varieties, and these polymorphisms are different between *rhg1-a* and *rhg1-b*

(Cook et al., 2014; Bayless et al., 2016). Additional evidence suggests that there are transcriptional and methylation differences between *rhg1-a* and *rhg1-b*, although these differences are less well characterized (Cook et al., 2014). In Chapter 4, we characterize a novel *Rhg1* haplotype found in the soybean crop wild relative, *Glycine soja*.

1.4 Vesicle Trafficking: SNAPs, SNAREs, and NSFs

Vesicle trafficking is an essential process to eukaryotic organisms, moving membrane-associated or vesicle-enclosed cargo between different cellular compartments. Core components of the vesicle trafficking system include SNAREs (soluble NSF attachment protein receptors), NSF (N-ethylmaleimide sensitive factor) and SNAP (soluble NSF attachment protein). During vesicle fusion with a target membrane, SNAREs on opposing membranes draw the two membranes together in an energetically favorable manner, forming a stable SNARE complex (Jahn and Scheller, 2006). This SNARE complex needs to be disassembled by SNAP and NSF in an ATP-dependent manner to allow for future rounds of vesicle fusion to occur (Jahn and Scheller, 2006). Additional components of the vesicle fusion machinery include the membrane tethering complex, sec/munc18 family proteins, Rab GTPases, as well as a panoply of vesicle-coat proteins (COPI, COPII, clathrin), which together serve to mediate initial fusion of opposed membranes through their SNAREs or mark where the vesicles are going (Wickner and Schekman, 2008; Inada and Ueda, 2014; Baker and Hughson, 2016).

Given their central role in eukaryotic cell homeostasis, NSF and α-SNAP were cloned and characterized thirty years ago. Initial studies on these key proteins identified domains central to their function. NSF is a AAA+ (ATPase associated with various cellular activities) protein that has

three domains: N, D1, and D2 (Jahn and Shceller, 2006; Zhao and Brunger, 2016). The NSF N domain is thought to be required for NSF association with α-SNAP, as 20S complex (the complex of four SNARE, four α-SNAPs and six NSF subunits; Figure 3) formation is impaired in deletion mutants (Nagiec et al., 1995; Yu et al., 1999). The D1 domain has high ATPase activity and is thought to drive the disassembly of SNARE bundles through the action of α-SNAP (Nagiec et al., 1995). The D2 domain has low ATPase activity, but is thought to mediate formation of the NSF hexamer, required for 20S assembly (Nagiec et al., 1995). α-SNAP is an adapter protein thought to both stimulate NSF ATPase activity as well as provide the anchoring force to allow disassembly of SNARE bundles. Mutational analyses identified conserved features of α-SNAP that allow it to bind promiscuously to different SNAREs through a concave face (Rice and Brunger, 1997; Marz et al., 2003). Specifically, an apolar loop at the N-terminus seems to play a role in SNARE binding (Zick et al., 2015). The α-SNAP C-terminus is highly conserved, and it is thought to mediate interactions with NSF (Barnard et al., 1996; Barnard et al., 1997; Rice and Brunger, 1999; Zhao et al., 2015). Particularly, a penultimate leucine is thought to stimulate the ATPase activity of NSF through interaction with D1 (Barnard et al., 1996; Barnard et al., 1997). These predictions have been robustly validated in more recent work, for example through cryo-EM and kinetic structure-function studies (Zhao et al., 2015; Zhou et al., 2015; White et al., 2018). The activities of NSF and α-SNAP are coordinately regulated by the availability of each to form 20S complexes as well as through posttranslational modifications such as phosphorylation and nitrosylation, which reduce 20S complex formation (Hirling and Scheller, 1996; Zhao et al., 2007; Barszczewski et al., 2008).

In addition to SNARE disassembly to promote vesicle trafficking, there is evidence that α-SNAP and NSF play additional roles throughout the cell. There is some evidence that α -SNAPs promote or inhibit SNARE bundle assembly (Barszczewski et al., 2008; Choi et al., 2018; Jun and Wickner, 2019). α-SNAP also has been associated with processes as varied as calcium sensing, glutamate sensing, cell plate deposition and apoptotic cell death (Hanley et al., 2002; Rancour et al., 2002; Naydenov et al., 2011; Miao et al., 2013). Interestingly, some of these α-SNAP- dependent processes seems not to require the activity of NSF. It bears mentioning that in addition to α-SNAP, plants and higher eukaryotes have two additional isoforms of SNAP: β-SNAP and γ-SNAP. Briefly, whereas β-SNAP seems to be brain specific, and so far, lacking in plants, γ-SNAP is ubiquitous (Whiteheart et al.,1993) but relatively un-studied. Furthermore, the different SNAP isoforms seem to have undergone subfunctionalization. This is best illustrated by a recent publication that shows γ-SNAP specifically interacts with mitochondrial tethering proteins and particular SNARE complexes, in spite of sharing an overall similar fold with α-SNAP (Chen et al., 2001; Bitto et al., 2007; Inoue et al., 2015).

The polyploid nature of soybean allows for further opportunities of subfunctionalization. Whereas most animals only encode a single α-SNAP and NSF protein (and one β- and γ-SNAP), the soybean genome encodes five α-SNAPs, two γ-SNAPs and two NSFs (Schmutz et al., 2010). It is worth mentioning that not all of these paralogs are equally expressed; rather they are expressed at varying degrees and in response to different signals, with some apparently having becoming pseudogenes. Apparently, a complex interplay between these partially redundant but subfunctionalized plant genes allows for the multifaceted nature of vesicle trafficking under homoeostatic conditions and during disease situations (discussed below).

Vesicle trafficking has been most prominently associated with plant disease and disease resistance with respect to extracellular delivery of antimicrobial compounds and small RNAs, cycling of the flagellin receptor FLS2, and in cytoskeletal rearrangements at sites of pathogen attack (Hoefle et al., 2008; Frey and Robatzek, 2009; Cai et al., 2019; Ekanyake et al., 2019). In the context of SCN, the *Rhg1* locus has an α-SNAP that has amino acid polymorphisms at the extreme C-terminus in resistant haplotypes (Cook et al., 2014; Bayless et al., 2016;). The resistant-type *Rhg1*-encoded α-SNAPs are unable to strongly interact with NSF, their expression disrupts normal vesicle trafficking, and these α -SNAPs accumulate within the syncytium that forms the biotrophic feeding site for cyst nematodes, suggesting disruption of susceptibility-associated vesicle trafficking in SCN resistance (Bayless et al., 2016; Bayless et al., 2019 [Chapter 3]). Membrane trafficking pathways present many potential targets for pathogens to modulate (Inada and Ueda, 2014). Recent work in elucidating effector families (proteins putatively delivered into the plant host cell) found that SNARE like proteins (SLPs) constitute one of the largest and most diverse groups of effectors in SCN (Gardener et al., 2018). There are functional data suggesting that one SLP, HgSLP-1, might physically interact with the α-SNAP from *Rhg1* (Bekal et al., 2015). Additional evidences suggest an important role of plant-encoded t-SNAREs in SCN resistance (Dong et al., 2020). However, the mechanisms underlying resistance or susceptibility outcomes, as well as whether there are additional effectors either from the SLP family or elsewhere, or perhaps even plant-encoded proteins, that interact with vesicle trafficking machinery, remains an open question. In Chapter 2, we clone and characterize a novel NSF allele that co-segregates with *Rhg1* and is necessary for viability in these varieties.

1.5 Long Terminal Repeat Retrotransposons in Plants

Retrotransposons of the long terminal repeat (LTR-RT) class are highly abundant within plant genomes (Schulman, 2013). Most LTR-RTs fall into two different families: *copia* and *gypsy*. The family an LTR-RT is classified as depends on gene order (Havecker et al., 2004; Zhao and Ma, 2013). Both types of families have long terminal repeats at both the 5' and 3' end that are capable of initiating transcription of the internal genes (Schulman, 2013). Between the terminal repeats is a polyprotein open reading frame encoding Gag, reverse transcriptase, integrase, and protease domains that are necessary for replication and transposition (Shulman, 2013). This polyprotein may then either be cleaved into individual proteins or a frameshift may be introduced to ultimately lead to independent proteins (Gao, 2003). However, there is diversity within the LTR-RT classes. Some LTR-RTs may have additional domains or additional gene fusions, and some may lack structural genes entirely, in the case of LARDs (large retrotransposon derivative) and TRIMs (terminal-repeat retrotransposons in miniature; Havecker et al., 2004). Transposition tends to favor integration into euchromatic regions proximal to host genes (Kidwell and Lisch, 2001; Lisch and Bennetzen, 2011).

Since their discovery over seventy years ago, transposons have been associated with modulating host phenotype (McClintock, 1950). Retrotransposons are thought to modulate host phenotypes in large part through their restructuring of genome architecture. Specifically, transposon insertion into genomes can impact genomes most dramatically by inactivating genes through insertion into or excision from gene bodies (Kidewell and Lisch, 2001; Mita and Boeke, 2016; Negi et al., 2016; Galindo-Gonzalez et al., 2017). More subtle changes include ways that transposons might structure genome architecture through promoting recombination

between regions of the genome that have LTR-RTs, leading to translocation, duplication, or deletion (Kidewell and Lisch, 2001; Mita and Boeke, 2016; Negi et al., 2016; Galindo-Gonzalez et al., 2017). LTR-RTs may also be located in introns. In this case, there is the possibility for promoting exon shuffling or creating splice variants through nascent splice sites within the LTR-RT (Kidewell and Lisch, 2001). Plant responses to transposable elements within the genome also dictate phenotypes. Under most conditions, plants have been selected to maintain low expression of transposons through methylation and RNAi mechanisms. However, when transposons are located proximal to genes, there can be additional silencing of these nearby genes (Kidwell and Lisch, 2001; Lisch and Bennetzen, 2011; Zhao and Ma, 2013; Kim and Zilberman, 2014; Mita and Boeke, 2016; Negi et al., 2016; Galindo-Gonzalez et al., 2017). Finally, because LTR-RTs have sequences within the LTRs that are sufficient to promote expression, nearby genes may have expression patterns changed to match those of the LTR-RT instead of the native promoter (Kidwell et al., 2001; Schulman et al., 2013; Zhao and Ma, 2013; Cavrak et al., 2014; Mita and Boeke, 2016; Negi et al., 2016). Through these mechanisms, differences in numerous traits have been reported such as flowering time, trichome development, and fruit size (Liu et al., 2004; Xiao et al., 2008; Ding et al., 2015).

Retroelements have also been implicated in disease resistance through changing promoter-proximal elements and choice of polyadenylation sites. Briefly, a promoter-proximal LTR-RT in the polymorphic rice *Pit* gene has been associated with higher levels of *Pit* expression and concomitant increase in rice blast resistance. Interestingly, chimera constructs suggested that this LTR-RT insertion was sufficient for improved rice blast resistance even when cloned into a susceptible rice variety (Hayashi and Yoshida, 2009). An additional example comes from

the *Arabidopsis RPP7* gene, which confers resistance to downy mildew. Here, the insertion of an LTR-RT was able to suppress choice of a premature polyadenylation site, increasing expression of full-length RPP7 protein (Tsuchiya and Eulgem, 2013). Genome architecture through histone methylation was implicated in this choice modification. In Chapter 3, we describe a retrotransposon found in *rhg1-a* type *G. max* varieties.

1.6 *Glycine soja*: Overview and use as a crop wild relative

Like many important crop species, soybean is primarily self-fertilizing. The resultingly high extent of homozygosity has resulted in the purging of deleterious alleles and contributed to the ease of breeding elite varieties used in commercial growing systems today. However, these same qualities have resulted in the purging of allelic diversity at other loci linked to the aforementioned deleterious alleles. Additionally, domestication and artificial selection from a small founding population have created genetic bottlenecks, leading to the more genetically homogenous cultivated species we know today (Halliburton, 2004; Liu et al., 2020). In fact, for soybean, it is estimated that the domestication bottleneck resulted in a nearly 80% decline in rare genes, with nearly 60% of retained genes exhibiting significant changes in allele frequency (Hyten et al., 2006). This loss of allelic diversity has profoundly affected plant phenotypes, ranging from disease resistance to abiotic stress tolerance (see below).

Because of the above, wild relatives of crop species represent an appealing choice to source novel traits of agronomic interest. For soybean, *G. soja* is a wild relative that remains sexually compatible with domesticated soybean, easing its use as a source for rare alleles that may aid crop improvement. Indeed, traits associated with improved yield, salt tolerance,

alkaline salt tolerance, and resistance to soybean aphid, foxglove aphid and SCN have all been found in *G. soja* (Wang et al., 2001; Concibido et al., 2003; Kabelka et al., 2005; Kabelka et al., 2006; Lee et al., 2009; Hesler, 2013; Kim and Diers, 2013; Lee et al., 2015; Yu and Diers, 2017). For SCN in particular, *G. soja* PI 468916 has been found to contain two loci that when introgressed into *Rhg1*-containing soybean provide additional resistance (Brzostowski and Diers, 2017). Of note, one of the genes in one of the loci encodes a γ-SNAP. Additional work has also suggested that complex regulatory networks underly SCN resistance in *G. soja* (Zhang et al., 2017). Recent sequencing and re-sequencing efforts of *G. soja* will further assist in gene discovery. In Chapter 4, we describe highly SCN-resistant *G. soja* that have a novel *Rhg1* and α-SNAP allele.

In sum, this thesis seeks to explore important loci that contribute to resistance in domesticated soybean (Chapters 2 and 3). We further look to elucidate the evolution of *Rhg1* by looking at the crop wild relative of soybean, *G. soja* (Chapter 4). In appendices, we describe methods useful for investigating the interaction between *Rhg1*-encoded proteins and other *G. max* proteins or SCN-encoded effector molecules (Appendices 1 and 4). We additionally speculate that *Rhg1*-encoded proteins may have broader utility in conferring disease resistance beyond the native SCN-soybean pathosystem (Appendices 2 and 3). We end by looking at *Rhg1* in other species, including common bean (*Phaseolus vulgaris*) and perennial *Glycine* species (Appendix 5).

1.7 References

Acharya, K., Tande, C. and Byamukama, E. (2016). Determination of Heterodera glycines Virulence Phenotypes Occurring in South Dakota. *Plant Disease*, 100(11), pp.2281–2286.

Allen, T.W., Bradley, C.A., Sisson, A.J., Byamukama, E., Chilvers, M.I., Coker, C.M., Collins, A.A., Damicone, J.P., Dorrance, A.E., Dufault, N.S., Esker, P.D., Faske, T.R., Giesler, L.J., Grybauskas, A.P., Hershman, D.E., Hollier, C.A., Isakeit, T., Jardine, D.J., Kelly, H.M., Kemerait, R.C., Kleczewski, N.M., Koenning, S.R., Kurle, J.E., Malvick, D.K., Markell, S.G., Mehl, H.L., Mueller, D.S., Mueller, J.D., Mulrooney, R.P., Nelson, B.D., Newman, M.A., Osborne, L., Overstreet, C., Padgett, G.B., Phipps, P.M., Price, P.P., Sikora, E.J., Smith, D.L., Spurlock, T.N., Tande, C.A., Tenuta, A.U., Wise, K.A. and Wrather, J.A. (2017). Soybean Yield Loss Estimates Due to Diseases in the United States and Ontario, Canada, from 2010 to 2014. *Plant Health Progress*, 18(1), pp.19–27.

de Almeida Engler, J., De Vleesschauwer, V., Burssens, S., Celenza, J.L., Inzé, D., Van Montagu, M., Engler, G. and Gheysen, G. (1999). Molecular Markers and Cell Cycle Inhibitors Show the Importance of Cell Cycle Progression in Nematode-Induced Galls and Syncytia. *The Plant Cell*, 11(5), pp.793–807.

de Almeida Engler, J. and Gheysen, G. (2013). Nematode-Induced Endoreduplication in Plant Host Cells: Why and How? *Molecular Plant-Microbe Interactions®*, 26(1), pp.17–24.

Anon, (n.d.). *http://soystats.com/*. [online] Available at: http://soystats.com [Accessed 2 Aug. 2020].

Baker, R.W. and Hughson, F.M. (2016). Chaperoning SNARE assembly and disassembly. *Nature reviews. Molecular cell biology*, [online] 17(8), pp.465–479. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5471617/ [Accessed 27 Aug. 2020].

Bao, Y., Chen, S., Vetsch, J. and Randall, G. (2013). Soybean Yield and Heterodera glycines Response to Liquid Swine Manure in Nematode Suppressive Soil and Conductive Soil. *The Journal of Nematology*, 45(1), pp.21–29.

Barnard, R.J., Morgan, A. and Burgoyne, R.D. (1996). Domains of alpha-SNAP required for the stimulation of exocytosis and for N-ethylmaleimide-sensitive fusion protein (NSF) binding and activation. *Molecular Biology of the Cell*, 7(5), pp.693–701.

Barnard, R.J.O., Morgan, A. and Burgoyne, R.D. (1997). Stimulation of NSF ATPase Activity by α-SNAP Is Required for SNARE Complex Disassembly and Exocytosis. *Journal of Cell Biology*, 139(4), pp.875–883.

Barszczewski, M., Chua, J.J., Stein, A., Winter, U., Heintzmann, R., Zilly, F.E., Fasshauer, D., Lang, T. and Jahn, R. (2008). A Novel Site of Action for α-SNAP in the SNARE Conformational Cycle Controlling Membrane Fusion. *Molecular Biology of the Cell*, 19(3), pp.776–784.

Bayless, A.M., Smith, J.M., Song, J., McMinn, P.H., Teillet, A., August, B.K. and Bent, A.F. (2016). Disease resistance through impairment of α-SNAP–NSF interaction and vesicular trafficking by soybean Rhg1. *Proceedings of the National Academy of Sciences*, 113(47), pp.E7375–E7382.

Bayless, A.M., Zapotocny, R.W., Han, S., Grunwald, D.J., Amundson, K.K. and Bent, A.F. (2019). The rhg1-a (Rhg1 low-copy) nematode resistance source harbors a copia-family retrotransposon within the Rhg1-encoded α-SNAP gene. *Plant Direct*, 3(8), p.e00164.

Bekal, S., Domier, L.L., Gonfa, B., Lakhssassi, N., Meksem, K. and Lambert, K.N. (2015). A SNARE-Like Protein and Biotin Are Implicated in Soybean Cyst Nematode Virulence. *PLOS ONE*, 10(12), p.e0145601.

Bekal, S., Niblack, T.L. and Lambert, K.N. (2003). A Chorismate Mutase from the Soybean Cyst Nematode Heterodera glycines Shows Polymorphisms that Correlate with Virulence. *Molecular Plant-Microbe Interactions®*, 16(5), pp.439–446.

Bitto, E., Bingman, C.A., Kondrashov, D.A., McCoy, J.G., Bannen, R.M., Wesenberg, G.E. and Phillips, G.N. (2007). Structure and dynamics of γ-SNAP: Insight into flexibility of proteins from the SNAP family. *Proteins: Structure, Function, and Bioinformatics*, 70(1), pp.93–104.

Brucker, E., Carlson, S., Wright, E., Niblack, T. and Diers, B. (2005). Rhg1 alleles from soybean PI 437654 and PI 88788 respond differentially to isolates of Heterodera glycines in the greenhouse. *Theoretical and Applied Genetics*, 111(1), pp.44–49.

Brzostowski, L.F. and Diers, B.W. (2017). Pyramiding of Alleles from Multiple Sources Increases the Resistance of Soybean to Highly Virulent Soybean Cyst Nematode Isolates. *Crop Science*, 57(6), pp.2932–2941.

Cai, Q., He, B. and Jin, H. (2019). A safe ride in extracellular vesicles – small RNA trafficking between plant hosts and pathogens. *Current Opinion in Plant Biology*, [online] 52, pp.140–148. Available at: https://www.sciencedirect.com/science/article/pii/S1369526618301560?via%3Dihub.

Cavrak, V.V., Lettner, N., Jamge, S., Kosarewicz, A., Bayer, L.M. and Mittelsten Scheid, O. (2014). How a Retrotransposon Exploits the Plant's Heat Stress Response for Its Activation. *PLoS Genetics*, 10(1), p.e1004115.

Chen, D., Xu, W., He, P., Medrano, E.E. and Whiteheart, S.W. (2001). Gaf-1, a γ-SNAP-binding Protein Associated with the Mitochondria. *Journal of Biological Chemistry*, 276(16), pp.13127–13135.

Chen, S., Porter, P.M., Reese, C.D. and Stienstra, W.C. (2001b). Crop Sequence Effects on Soybean Cyst Nematode and Soybean and Corn Yields. *Crop Science*, 41(6), pp.1843–1849.

Choi, U.B., Zhao, M., White, K.I., Pfuetzner, R.A., Esquivies, L., Zhou, Q. and Brunger, A.T. (2018). NSF-mediated disassembly of on- and off-pathway SNARE complexes and inhibition by complexin. *eLife*, 7.

Colgrove, A. and Niblack, T. (2008). Correlation of female indices from virulence assays on inbred lines and field populations of Heterodera glycines. *The Journal of Nematology*, 40(1), pp.39–45.

Concibido, V., La Vallee, B., Mclaird, P., Pineda, N., Meyer, J., Hummel, L., Yang, J., Wu, K. and Delannay, X. (2003). Introgression of a quantitative trait locus for yield from Glycine soja into commercial soybean cultivars. *Theoretical and Applied Genetics*, 106(4), pp.575–582.

Cook, D.E., Bayless, A.M., Wang, K., Guo, X., Song, Q., Jiang, J. and Bent, A.F. (2014). Distinct Copy Number, Coding Sequence, and Locus Methylation Patterns Underlie Rhg1-Mediated Soybean Resistance to Soybean Cyst Nematode. *Plant Physiology*, 165(2), pp.630–647.

Cook, D.E., Lee, T.G., Guo, X., Melito, S., Wang, K., Bayless, A.M., Wang, J., Hughes, T.J., Willis, D.K., Clemente, T.E., Diers, B.W., Jiang, J., Hudson, M.E. and Bent, A.F. (2012). Copy Number Variation of Multiple Genes at Rhg1 Mediates Nematode Resistance in Soybean. *Science*, 338(6111), pp.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, A.H. and 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(1), pp.143–154.

Dong, J., Zielinski, R.E. and Hudson, M.E. (2020). t-SNAREs Bind the Rhg1 α-SNAP and Mediate Soybean Cyst Nematode Resistance. *The Plant Journal*.

Ekanayake, G., LaMontagne, E.D. and Heese, A. (2019). Never Walk Alone: Clathrin-Coated Vesicle (CCV) Components in Plant Immunity. *Annual Review of Phytopathology*, 57(1), pp.387–409.

Eppo.int. (2020). *home page*. [online] Available at: https://www.eppo.int [Accessed 17 Oct. 2020].

Eves-van den Akker, S. and Birch, P.R.J. (2016). Opening the Effector Protein Toolbox for Plant–Parasitic Cyst Nematode Interactions. *Molecular Plant*, 9(11), pp.1451–1453.

Faghihi, J., Donald, P.A., Noel, G., Welacky, T.W. and Ferris, V.R. (2010). Soybean Resistance to Field Populations of Heterodera glycines in Selected Geographic Areas. *Plant Health Progress*, 11(1), p.19.

Fao.org. (2019). *FAOSTAT*. [online] Available at: http://www.fao.org/faostat/en/#data/QC.

Frey, N.F. dit and Robatzek, S. (2009). Trafficking vesicles: pro or contra pathogens? *Current Opinion in Plant Biology*, 12(4), pp.437–443.

Galindo-González, L., Mhiri, C., Deyholos, M.K. and Grandbastien, M.-A. (2017). LTR-retrotransposons in plants: Engines of evolution. *Gene*, 626, pp.14–25.

Gao, X. (2003). Translational recoding signals between gag and pol in diverse LTR retrotransposons. *RNA*, 9(12), pp.1422–1430.

Gardner, M., Dhroso, A., Johnson, N., Davis, E.L., Baum, T.J., Korkin, D. and Mitchum, M.G. (2018). Novel global effector mining from the transcriptome of early life stages of the soybean cyst nematode Heterodera glycines. *Scientific Reports*, 8(1).

Gaspar, A.P., Marburger, D.A., Mourtzinis, S. and Conley, S.P. (2014). Soybean Seed Yield Response to Multiple Seed Treatment Components across Diverse Environments. *Agronomy Journal*, 106(6), pp.1955–1962.

Gheysen, G. and Mitchum, M.G. (2011). How nematodes manipulate plant development pathways for infection. *Current Opinion in Plant Biology*, 14(4), pp.415–421.

Goverse, A., de Almeida Engler, J., Verhees, J., van der Krol, S., Helder, J. and Gheysen, G. (2000). Cell Cycle Activation by Plant Parasitic Nematodes. *Plant Molecular Biology*, 43(5/6), pp.747–761.

Grabau, Z.J. and Chen, S. (2016). Determining the Role of Plant-Parasitic Nematodes in the Corn-Soybean Crop Rotation Yield Effect Using Nematicide Application: II. Soybean. *Agronomy Journal*, 108(3), pp.1168–1179.

Grunewald, W., Cannoot, B., Friml, J. and Gheysen, G. (2009). Parasitic Nematodes Modulate PIN-Mediated Auxin Transport to Facilitate Infection. *PLoS Pathogens*, 5(1), p.e1000266.

Haarith, D., Kim, D., Strom, N.B., Chen, S. and Bushley, K.E. (2020). In Vitro Screening of a Culturable Soybean Cyst Nematode Cyst Mycobiome for Potential Biological Control Agents and Biopesticides. *Phytopathology®*, 110(8), pp.1388–1397.

Halliburton, R. (2004). *Introduction to population genetics*. Upper Saddle River, Nj: Pearson/Prentice Hall.

Hamamouch, N., Li, C., Hewezi, T., Baum, T.J., Mitchum, M.G., Hussey, R.S., Vodkin, L.O. and Davis, E.L. (2012). The interaction of the novel 30C02 cyst nematode effector protein with a plant β-1,3-endoglucanase may suppress host defence to promote parasitism. *Journal of Experimental Botany*, 63(10), pp.3683–3695.

Hanley, J.G., Khatri, L., Hanson, P.I. and Ziff, E.B. (2002). NSF ATPase and α-/β-SNAPs Disassemble the AMPA Receptor-PICK1 Complex. *Neuron*, 34(1), pp.53–67.

Hartman, G.L., West, E.D. and Herman, T.K. (2011). Crops that feed the World 2. Soybean—worldwide production, use, and constraints caused by pathogens and pests. *Food Security*, 3(1), pp.5–17.

Havecker, E.R., Gao, X. and Voytas, D.F. (2004). The Diversity of LTR Retrotransposons. *Genome Biology*, 5(6), p.225.

Hayashi, K. and Yoshida, H. (2009). Refunctionalization of the ancient rice blast disease resistance genePitby the recruitment of a retrotransposon as a promoter. *The Plant Journal*, 57(3), pp.413–425.

Hershman, D.E., Heinz, R.D. and Kennedy, B.S. (2008). Soybean Cyst Nematode, Heterodera glycines, Populations Adapting to Resistant Soybean Cultivars in Kentucky. *Plant Disease*, 92(10), pp.1475–1475.

Hesler, L.S. (2013). Resistance to soybean aphid among wild soybean lines under controlled conditions. *Crop Protection*, 53, pp.139–146.

Hewezi, T. and Baum, T.J. (2013). Manipulation of Plant Cells by Cyst and Root-Knot Nematode Effectors. *Molecular Plant-Microbe Interactions®*, 26(1), pp.9–16.

Hewezi, T., Howe, P., Maier, T.R. and Baum, T.J. (2008). Arabidopsis Small RNAs and Their Targets During Cyst Nematode Parasitism. *Molecular Plant-Microbe Interactions®*, 21(12), pp.1622–1634.

Hewezi, T., Juvale, P.S., Piya, S., Maier, T.R., Rambani, A., Rice, J.H., Mitchum, M.G., Davis, E.L., Hussey, R.S. and Baum, T.J. (2015). The Cyst Nematode Effector Protein 10A07 Targets and Recruits Host Posttranslational Machinery to Mediate Its Nuclear Trafficking and to Promote Parasitism in Arabidopsis. *The Plant Cell*, 27(3), pp.891–907.

Hirling, H. and Scheller, R.H. (1996). Phosphorylation of synaptic vesicle proteins: modulation of the alpha SNAP interaction with the core complex. *Proceedings of the National Academy of Sciences*, 93(21), pp.11945–11949.

Hoefle, C. and Hückelhoven, R. (2008). Enemy at the gates: traffic at the plant cell pathogen interface. *Cellular Microbiology*, 10(12), pp.2400–2407.

Hyten, D.L., Song, Q., Zhu, Y., Choi, I.-Y., Nelson, R.L., Costa, J.M., Specht, J.E., Shoemaker, R.C. and Cregan, P.B. (2006). Impacts of genetic bottlenecks on soybean genome diversity. *Proceedings of the National Academy of Sciences*, 103(45), pp.16666–16671.

Inada, N. and Ueda, T. (2014). Membrane Trafficking Pathways and their Roles in Plant–Microbe Interactions. *Plant and Cell Physiology*, 55(4), pp.672–686.

Inoue, H., Matsuzaki, Y., Tanaka, A., Hosoi, K., Ichimura, K., Arasaki, K., Wakana, Y., Asano, K., Tanaka, M., Okuzaki, D., Yamamoto, A., Tani, K. and Tagaya, M. (2015). -SNAP stimulates

disassembly of endosomal SNARE complexes and regulates endocytic trafficking pathways. *Journal of Cell Science*, 128(15), pp.2781–2794.

Jahn, R. and Scheller, R.H. (2006). SNAREs — engines for membrane fusion. *Nature Reviews Molecular Cell Biology*, [online] 7(9), pp.631–643. Available at: https://www.nature.com/articles/nrm2002 [Accessed 1 May 2020].

Jones, J.T., Haegeman, A., Danchin, E.G.J., Gaur, H.S., Helder, J., Jones, M.G.K., Kikuchi, T., Manzanilla-López, R., Palomares-Rius, J.E., Wesemael, W.M.L. and Perry, R.N. (2013). Top 10 plant-parasitic nematodes in molecular plant pathology. *Molecular Plant Pathology*, 14(9), pp.946–961.

Jun, Y. and Wickner, W. (2019). Sec17 (α-SNAP) and Sec18 (NSF) restrict membrane fusion to R-SNAREs, Q-SNAREs, and SM proteins from identical compartments. *Proceedings of the National Academy of Sciences*, 116(47), pp.23573–23581.

Kabelka, E.A., Carlson, S.R. and Diers, B.W. (2005). Localization of Two Loci that Confer Resistance to Soybean Cyst Nematode from Glycine soja PI 468916. *Crop Science*, 45(6), pp.2473–2481.

Kabelka, E.A., Carlson, S.R. and Diers, B.W. (2006). Glycine soja PI 468916 SCN Resistance Loci's Associated Effects on Soybean Seed Yield and Other Agronomic Traits. *Crop Science*, 46(2), pp.622–629.

Kidwell, M.G. and Lisch, D.R. (2001). PERSPECTIVE: TRANSPOSABLE ELEMENTS, PARASITIC DNA, AND GENOME EVOLUTION. *Evolution*, 55(1), pp.1–24.

Kim, M. and Diers, B.W. (2013). Fine Mapping of the SCN Resistance QTL and from PI 468916. *Crop Science*, 53(3), p.775.

Kim, M., Hyten, D.L., Bent, A.F. and Diers, B.W. (2010). Fine Mapping of the SCN Resistance Locus rhg1-b from PI 88788. *The Plant Genome Journal*, 3(2), p.81.

Kim, M.Y., Lee, S., Van, K., Kim, T.-H., Jeong, S.-C., Choi, I.-Y., Kim, D.-S., Lee, Y.-S., Park, D., Ma, J., Kim, W.-Y., Kim, B.-C., Park, S., Lee, K.-A., Kim, D.H., Kim, K.H., Shin, J.H., Jang, Y.E., Kim, K.D., Liu, W.X., Chaisan, T., Kang, Y.J., Lee, Y.-H., Kim, K.-H., Moon, J.-K., Schmutz, J., Jackson, S.A., Bhak, J. and Lee, S.-H. (2010b). Whole-genome sequencing and intensive analysis of the undomesticated soybean (Glycine soja Sieb. and Zucc.) genome. *Proceedings of the National Academy of Sciences*, 107(51), pp.22032–22037.

Kim, M.Y. and Zilberman, D. (2014). DNA methylation as a system of plant genomic immunity. *Trends in Plant Science*, 19(5), pp.320–326.

Koenning, S.R. and Wrather, J.A. (2010). Suppression of Soybean Yield Potential in the Continental United States by Plant Diseases from 2006 to 2009. *Plant Health Progress*, 11(1), p.5.

Kofsky, J., Zhang, H. and Song, B.-H. (2018). The Untapped Genetic Reservoir: The Past, Current, and Future Applications of the Wild Soybean (Glycine soja). *Frontiers in Plant Science*, 9.

Kyndt, T., Vieira, P., Gheysen, G. and de Almeida-Engler, J. (2013). Nematode feeding sites: unique organs in plant roots. *Planta*, 238(5), pp.807–818.

Lahm, G.P., Desaeger, J., Smith, B.K., Pahutski, T.F., Rivera, M.A., Meloro, T., Kucharczyk, R., Lett, R.M., Daly, A., Smith, B.T., Cordova, D., Thoden, T. and Wiles, J.A. (2017). The discovery of fluazaindolizine: A new product for the control of plant parasitic nematodes. *Bioorganic & Medicinal Chemistry Letters*, 27(7), pp.1572–1575.

Lauritis, J., Rebois, R. and Graney, L. (1983). Development of Heterodera glycines Ichinohe on Soybean, Glycine max (L.) Merr., under Gnotobiotic Conditions. *The Journal of Nematology*, 15(2), pp.272–281.

Lee, J.-D., Shannon, J.G., Vuong, T.D. and Nguyen, H.T. (2009). Inheritance of Salt Tolerance in Wild Soybean (Glycine soja Sieb. and Zucc.) Accession PI483463. *Journal of Heredity*, 100(6), pp.798–801.

Lee, J.S., Yoo, M., Jung, J.K., Bilyeu, K.D., Lee, J.-D. and Kang, S. (2015). Detection of novel QTLs for foxglove aphid resistance in soybean. *Theoretical and Applied Genetics*, 128(8), pp.1481–1488.

Lee, T.G., Kumar, I., Diers, B.W. and Hudson, M.E. (2015b). Evolution and selection ofRhg1,a copy-number variant nematode-resistance locus. *Molecular Ecology*, 24(8), pp.1774–1791.

Li, J., Todd, T.C., Oakley, T.R., Lee, J. and Trick, H.N. (2010). Host-derived suppression of nematode reproductive and fitness genes decreases fecundity of Heterodera glycines Ichinohe. *Planta*, 232(3), pp.775–785.

Li, Y., Zhou, G., Ma, J., Jiang, W., Jin, L., Zhang, Z., Guo, Y., Zhang, J., Sui, Y., Zheng, L., Zhang, S., Zuo, Q., Shi, X., Li, Y., Zhang, W., Hu, Y., Kong, G., Hong, H., Tan, B., Song, J., Liu, Z., Wang, Y., Ruan, H., Yeung, C.K.L., Liu, J., Wang, H., Zhang, L., Guan, R., Wang, K., Li, W., Chen, S., Chang, R., Jiang, Z., Jackson, S.A., Li, R. and Qiu, L. (2014). De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nature Biotechnology*, [online] 32(10), pp.1045–1052. Available at: https://www.nature.com/articles/nbt.2979 [Accessed 27 Dec. 2019].

Lightfoot, D. and Meksem, K. (2001). *Isolated Polynucleotides and Polypeptides Relating to Loci Underlying Resistance to Soybean Cyst Nematode and Soybean Sudden Death Syndrome and Methods Employing the Same*. [online] Available at: https://patents.google.com/patent/CA2331674A1/en#citedBy.

Lisch, D. and Bennetzen, J.L. (2011). Transposable element origins of epigenetic gene regulation. *Current Opinion in Plant Biology*, 14(2), pp.156–161.

Liu, J., He, Y., Amasino, R. and Chen, X. (2004). siRNAs targeting an intronic transposon in the regulation of natural flowering behavior in Arabidopsis. *Genes & Development*, 18(23), pp.2873–2878.

Liu, Y., Du, H., Li, P., Shen, Y., Peng, H., Liu, S., Zhou, G.-A., Zhang, H., Liu, Z., Shi, M., Huang, X., Li, Y., Zhang, M., Wang, Z., Zhu, B., Han, B., Liang, C. and Tian, Z. (2020). Pan-Genome of Wild and Cultivated Soybeans. *Cell*, 182(1), pp.162-176.e13.

Marz, K.E., Lauer, J.M. and Hanson, P.I. (2003). Defining the SNARE Complex Binding Surface of α-SNAP. *Journal of Biological Chemistry*, 278(29), pp.27000–27008.

Masonbrink, R., Maier, T.R., Muppirala, U., Seetharam, A.S., Lord, E., Juvale, P.S., Schmutz, J., Johnson, N.T., Korkin, D., Mitchum, M.G., Mimee, B., den Akker, S.E., Hudson, M., Severin, A.J. and Baum, T.J. (2019). The genome of the soybean cyst nematode (Heterodera glycines) reveals complex patterns of duplications involved in the evolution of parasitism genes. *BMC Genomics*, 20(1).

McCarville, M.T., Marett, C.C., Mullaney, M.P., Gebhart, G.D. and Tylka, G.L. (2017). Increase in Soybean Cyst Nematode Virulence and Reproduction on Resistant Soybean Varieties in Iowa From 2001 to 2015 and the Effects on Soybean Yields. *Plant Health Progress*, 18(3), pp.146–155.

McClintock, B. (1950). The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences*, 36(6), pp.344–355.

Mejias, J., Truong, N.M., Abad, P., Favery, B. and Quentin, M. (2019). Plant Proteins and Processes Targeted by Parasitic Nematode Effectors. *Frontiers in Plant Science*, 10.

Melito, S., Heuberger, A.L., Cook, D., Diers, B.W., MacGuidwin, A.E. and Bent, A.F. (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(1), p.104.

Miao, Y., Miner, C., Zhang, L., Hanson, P.I., Dani, A. and Vig, M. (2013). An essential and NSF independent role for α-SNAP in store-operated calcium entry. *eLife*, 2.

Mita, P. and Boeke, J.D. (2016). How retrotransposons shape genome regulation. *Current Opinion in Genetics & Development*, 37, pp.90–100.

Mitchum, M.G. (2016). Soybean Resistance to the Soybean Cyst Nematode Heterodera glycines: An Update. *Phytopathology®*, 106(12), pp.1444–1450.

Mitchum, M.G., Hussey, R.S., Baum, T.J., Wang, X., Elling, A.A., Wubben, M. and Davis, E.L. (2013). Nematode effector proteins: an emerging paradigm of parasitism. *New Phytologist*, 199(4), pp.879–894.

Mitchum, M.G., Wrather, J.A., Heinz, R.D., Shannon, J.G. and Danekas, G. (2007). Variability in Distribution and Virulence Phenotypes of Heterodera glycines in Missouri During 2005. *Plant Disease*, 91(11), pp.1473–1476.

Nagiec, E.E., Bernstein, A. and Whiteheart, S.W. (1995). Each Domain of the N-Ethylmaleimide-sensitive Fusion Protein Contributes to Its Transport Activity. *Journal of Biological Chemistry*, 270(49), pp.29182–29188.

Naydenov, N.G., Harris, G., Brown, B., Schaefer, K.L., Das, S.K., Fisher, P.B. and Ivanov, A.I. (2011). Loss of Soluble N-Ethylmaleimide-sensitive Factor Attachment Protein α (αSNAP) Induces Epithelial Cell Apoptosis via Down-regulation of Bcl-2 Expression and Disruption of the Golgi. *Journal of Biological Chemistry*, 287(8), pp.5928–5941.

Negi, P., Rai, A.N. and Suprasanna, P. (2016). Moving through the Stressed Genome: Emerging Regulatory Roles for Transposons in Plant Stress Response. *Frontiers in Plant Science*, 7.

Niblack, T., Arelli, P., Noel, G., Opperman, C., Orf, J., Schmitt, D., Shannon, J. and Tylka, G. (2002). A Revised Classification Scheme for Genetically Diverse Populations of Heterodera glycines. *The Journal of Nematology*, 34(4), pp.279–288.

Niblack, T.L., Colgrove, A.L., Colgrove, K. and Bond, J.P. (2008). Shift in Virulence of Soybean Cyst Nematode is Associated with Use of Resistance from PI 88788. *Plant Health Progress*, 9(1), p.29.

Niblack, T.L., Lambert, K.N. and Tylka, G.L. (2006). A Model Plant Pathogen from the Kingdom Animalia: Heterodera glycines, the Soybean Cyst Nematode. *Annual Review of Phytopathology*, 44(1), pp.283–303.

Nicol, J.M., Turner, S.J., Coyne, D.L., Nijs, L. den, Hockland, S. and Maafi, Z.T. (2011). Current Nematode Threats to World Agriculture. *Genomics and Molecular Genetics of Plant-Nematode Interactions*, [online] pp.21–43. Available at: https://link.springer.com/chapter/10.1007%2F978-94-007-0434-3_2 [Accessed 8 Dec. 2019].

Peng, D.L., Peng, H., Wu, D.Q., Huang, W.K., Ye, W.X. and Cui, J.K. (2016). First Report of Soybean Cyst Nematode (Heterodera glycines) on Soybean from Gansu and Ningxia China. *Plant Disease*, 100(1), pp.229–229.

Qin, L., Kudla, U., Roze, E.H.A., Goverse, A., Popeijus, H., Nieuwland, J., Overmars, H., Jones, J.T., Schots, A., Smant, G., Bakker, J. and Helder, J. (2004). A nematode expansin acting on plants. *Nature*, 427(6969), pp.30–30.

Rancour, D.M., Dickey, C., Park, S. and Bednarek, S. (2002). Characterization of AtCDC48. Evidence for Multiple Membrane Fusion Mechanisms at the Plane of Cell Division in Plants. *PLANT PHYSIOLOGY*, 130(3), pp.1241–1253.

Rice, L.M. and Brunger, A.T. (1999). Crystal Structure of the Vesicular Transport Protein Sec17. *Molecular Cell*, 4(1), pp.85–95.

Rincker, K., Cary, T. and Diers, B.W. (2017). Impact of Soybean Cyst Nematode Resistance on Soybean Yield. *Crop Science*, 57(3), p.1373.

Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D.L., Song, Q., Thelen, J.J., Cheng, J., Xu, D., Hellsten, U., May, G.D., Yu, Y., Sakurai, T., Umezawa, T., Bhattacharyya, M.K., 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, X.-C., Shinozaki, K., Nguyen, H.T., Wing, R.A., Cregan, P., Specht, J., Grimwood, J., Rokhsar, D., Stacey, G., Shoemaker, R.C. and Jackson, S.A. (2010). Genome sequence of the palaeopolyploid soybean. *Nature*, 463(7278), pp.178–183.

Schulman, A.H. (2013). Retrotransposon replication in plants. *Current Opinion in Virology*, 3(6), pp.604–614.

Siddique, S., Radakovic, Z.S., De La Torre, C.M., Chronis, D., Novák, O., Ramireddy, E., Holbein, J., Matera, C., Hütten, M., Gutbrod, P., Anjam, M.S., Rozanska, E., Habash, S., Elashry, A., Sobczak, M., Kakimoto, T., Strnad, M., Schmülling, T., Mitchum, M.G. and Grundler, F.M.W. (2015). A parasitic nematode releases cytokinin that controls cell division and orchestrates feeding site formation in host plants. *Proceedings of the National Academy of Sciences of the United States of America*, [online] 112(41), pp.12669–74. Available at: https://www.ncbi.nlm.nih.gov/pubmed/26417108 [Accessed 28 Dec. 2019].

Sindhu, A.S., Maier, T.R., Mitchum, M.G., Hussey, R.S., Davis, E.L. and Baum, T.J. (2008). Effective and specific in planta RNAi in cyst nematodes: expression interference of four parasitism genes reduces parasitic success. *Journal of Experimental Botany*, 60(1), pp.315–324.

Smant, G., Stokkermans, J.P.W.G., Yan, Y., de Boer, J.M., Baum, T.J., Wang, X., Hussey, R.S., Gommers, F.J., Henrissat, B., Davis, E.L., Helder, J., Schots, A. and Bakker, J. (1998). Endogenous cellulases in animals: Isolation of -1,4-endoglucanase genes from two species of plant-parasitic cyst nematodes. *Proceedings of the National Academy of Sciences*, 95(9), pp.4906–4911.

Strom, N., Hu, W., Haarith, D., Chen, S. and Bushley, K. (2020). Corn and Soybean Host Root Endophytic Fungi with Toxicity Toward the Soybean Cyst Nematode. *Phytopathology®*, 110(3), pp.603–614.

Tsuchiya, T. and Eulgem, T. (2013). An alternative polyadenylation mechanism coopted to the Arabidopsis RPP7 gene through intronic retrotransposon domestication. *Proceedings of the National Academy of Sciences*, 110(37), pp.E3535–E3543.

Turner, S.J. and Rowe, J.A. (2006). Cyst nematodes. *Plant nematology*, pp.91–122.

Tylka, G (2016). *More SCN-Resistant Soybean Varieties Than Ever, but Diversity of Resistance is Lacking*. [online] www.plantpath.iastate.edu. Available at: https://www.plantpath.iastate.edu/news/more-scn-resistant-soybean-varieties-ever-diversity-resistance-lacking [Accessed 17 Oct. 2020].

Tylka, G.L. and Marett, C.C. (2017). Known Distribution of the Soybean Cyst Nematode, Heterodera glycines, in the United States and Canada, 1954 to 2017. *Plant Health Progress*, 18(3), pp.167–168.

Wang, D., Duan, Y.X., Wang, Y.Y., Zhu, X.F., Chen, L.J., Liu, X.Y. and Chen, J.S. (2015). First Report of Soybean Cyst Nematode, Heterodera glycines, on Soybean From Guangxi, Guizhou, and Jiangxi Provinces, China. *Plant Disease*, 99(6), pp.893–893.

Wang, J., Lee, C., Replogle, A., Joshi, S., Korkin, D., Hussey, R., Baum, T.J., Davis, E.L., Wang, X. and Mitchum, M.G. (2010). Dual roles for the variable domain in protein trafficking and host-specific recognition of Heterodera glycines CLE effector proteins. *New Phytologist*, 187(4), pp.1003–1017.

WANG, J., REPLOGLE, A., HUSSEY, R., BAUM, T., WANG, X., DAVIS, E.L. and MITCHUM, M.G. (2010b). Identification of potential host plant mimics of CLAVATA3/ESR (CLE)-like peptides from the plant-parasitic nematode Heterodera schachtii. *Molecular Plant Pathology*, 12(2), pp.177–186.

Wang, J., Yeckel, G., Kandoth, P.K., Wasala, L., Hussey, R.S., Davis, E.L., Baum, T.J. and Mitchum, M.G. (2020). Targeted suppression of soybean BAG6-induced cell death in yeast by soybean cyst nematode effectors. *Molecular Plant Pathology*, 21(9), pp.1227–1239.

Webb, D.M., Baltazar, B.M., Rao-Arelli, A.P., Schupp, J., Clayton, K., Keim, P. and Beavis, W.D. (1995). Genetic mapping of soybean cyst nematode race-3 resistance loci in the soybean PI 437.654. *Theoretical and Applied Genetics*, 91(4), pp.574–581.

White, K.I., Zhao, M., Choi, U.B., Pfuetzner, R.A. and Brunger, A.T. (2018). Structural principles of SNARE complex recognition by the AAA+ protein NSF. *eLife*, 7.

Whiteheart, S.W., Griff, I.C., Brunner, M., Clary, D.O., Mayer, T., Buhrow, S.A. and Rothman, J.E. (1993). SNAP family of NSF attachment proteins includes a brain-specific isoform. *Nature*, 362(6418), pp.353–355.

Wickner, W. and Schekman, R. (2008). Membrane fusion. *Nature Structural & Molecular Biology*, 15(n7), pp.658–664.

Winstead, N., Skotland, C. and Sasser, J. (1955). Soybean Cyst Nematode in North Carolina. *Plant Disease Reporter*, 39(1), pp.9–11.

Xiao, H., Jiang, N., Schaffner, E., Stockinger, E.J. and van der Knaap, E. (2008). A Retrotransposon-Mediated Gene Duplication Underlies Morphological Variation of Tomato Fruit. *Science*, 319(5869), pp.1527–1530.

Yu, N. and Diers, B.W. (2017). Fine mapping of the SCN resistance QTL cqSCN-006 and cqSCN-007 from Glycine soja PI 468916. *Euphytica*, 213(2).

Yu, R.C., Jahn, R. and Brunger, A.T. (1999). NSF N-Terminal Domain Crystal Structure. *Molecular Cell*, 4(1), pp.97–107.

Zhang, H., Kjemtrup-Lovelace, S., Li, C., Luo, Y., Chen, L.P. and Song, B.-H. (2017). Comparative RNA-Seq Analysis Uncovers a Complex Regulatory Network for Soybean Cyst Nematode Resistance in Wild Soybean (Glycine soja). *Scientific Reports*, 7(1).

Zhao, C., Slevin, J.T. and Whiteheart, S.W. (2007). Cellular functions of NSF: Not just SNAPs and SNAREs. *FEBS Letters*, 581(11), pp.2140–2149.

Zhao, M. and Brunger, A.T. (2016). Recent Advances in Deciphering the Structure and Molecular Mechanism of the AAA+ ATPase N-Ethylmaleimide-Sensitive Factor (NSF). *Journal of Molecular Biology*, 428(9), pp.1912–1926.

Zhao, M. and Ma, J. (2013). Co-evolution of plant LTR-retrotransposons and their host genomes. *Protein & Cell*, 4(7), pp.493–501.

Zhao, M., Wu, S., Zhou, Q., Vivona, S., Cipriano, D.J., Cheng, Y. and Brunger, A.T. (2015). Mechanistic insights into the recycling machine of the SNARE complex. *Nature*, 518(7537), pp.61–67.

Zhou, Q., Huang, X., Sun, S., Li, X., Wang, H.-W. and Sui, S.-F. (2015). Cryo-EM structure of SNAP-SNARE assembly in 20S particle. *Cell Research*, 25(5), pp.551–560.

Zhou, Y., Wang, Y., Zhu, X., Liu, R., Xiang, P., Chen, J., Liu, X., Duan, Y. and Chen, L. (2017). Management of the soybean cyst nematode Heterodera glycines with combinations of different rhizobacterial strains on soybean. *PLOS ONE*, 12(8), p.e0182654.

Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, W., Yu, Y., Shu, L., Zhao, Y., Ma, Y., Fang, C., Shen, Y., Liu, T., Li, C., Li, Q., Wu, M., Wang, M., Wu, Y., Dong, Y., Wan, W., Wang, X., Ding, Z., Gao, Y., Xiang, H., Zhu, B., Lee, S.-H., Wang, W. and Tian, Z. (2015b). Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nature Biotechnology*, 33(4), pp.408–414.

Zick, M., Orr, A., Schwartz, M.L., Merz, A.J. and Wickner, W.T. (2015). Sec17 can trigger fusion of trans-SNARE paired membranes without Sec18. *Proceedings of the National Academy of Sciences*, 112(18), pp.E2290–E2297.

1.8 Figures

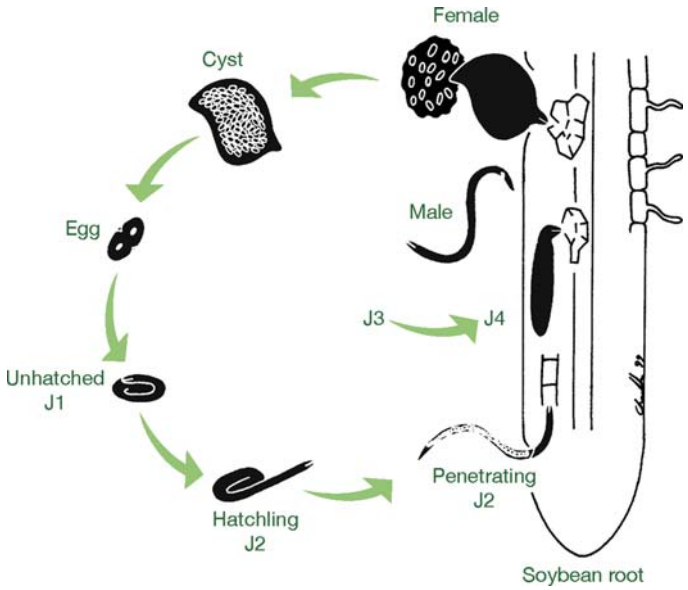


Fig. 1 Lifecycle of soybean cyst nematode (SCN). An unhatched J1 develops in the egg. The J1 undergoes one round of molting within the egg and emerges as a J2. The J2 penetrates host roots and undergoes two rounds of development (J3 and J4). The male regains vermiform shape and fertilizes a female. The fertilized female produces eggs both externally (as an egg mass) and internally fills with eggs. The female then develops a melanized, environmentally recalcitrant, and egg-filled cyst. From Niblack *et al.*, 2006.

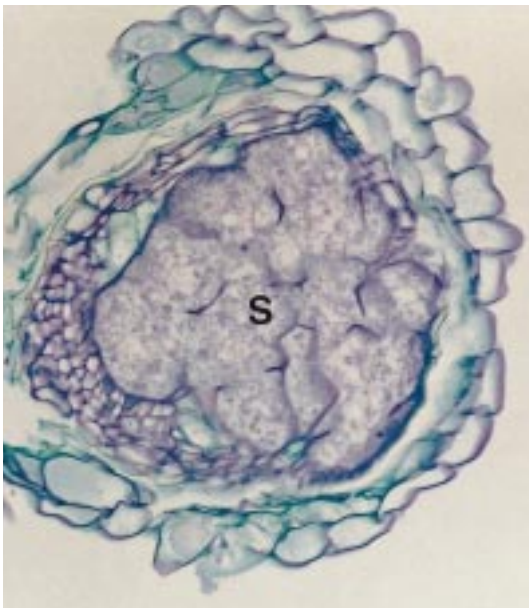


Fig. 2 Syncytium induced by the beet cyst nematode (*Heterodera schachtii*) on root cells of *Arabidopsis thaliana*. Cells stained with toluidine blue. Note loss of central vacuole and high density of cytosol in syncytial cell relative to surrounding cells. From Goverse *et al*., 2000.

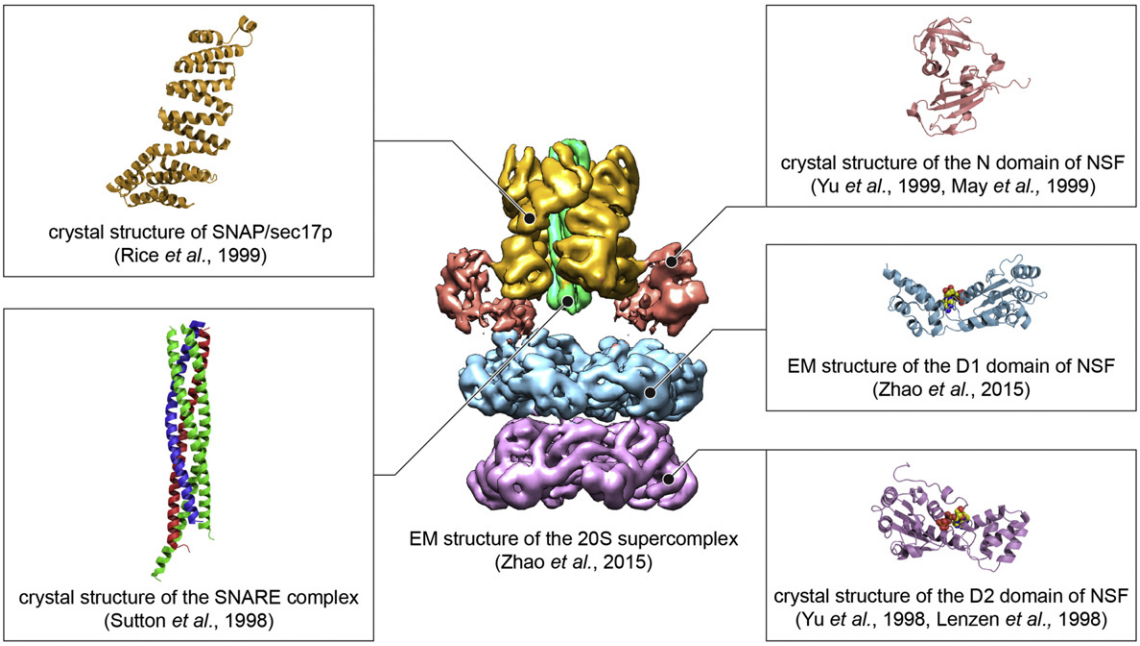


Figure 3 Cryo EM structure of the 20S supercomplex (center). Individual components of the 20S supercomplex (SNAP, SNARE, NSF) with dates of elucidation in boxes surrounding the EM structure. From Zhao and Brunger, 2016.

Chapter 2: An atypical NSF (N-ethylmaleimide Sensitive Factor) enables the viability of nematode-resistant *Rhg1* soybeans

Bayless, A.M., Zapotocny, R.W., Grunwald, D.J., Amundson, K.K., Diers, B.W. and Bent, A.F. (2018). An atypical N-ethylmaleimide sensitive factor enables the viability of nematode-resistant Rhg1 soybeans. *Proceedings of the National Academy of Sciences*, 115(19), pp.E4512–E4521.

Author contributions: I performed immunoblots on low-copy soybean accessions, assisted in binding assays, and genotyped the NAM soybean population.

Chapter 2: An atypical NSF (N-ethylmaleimide Sensitive Factor) enables the viability of nematode-resistant *Rhg1* soybeans

Adam M. Bayless[a], Ryan W. Zapotocny[a], Derrick J. Grunwald[a], Kaela Amundson, Brian W. Diers[b], Andrew F. Bent[a,1].

[a]Department of Plant Pathology, University of Wisconsin - Madison, Madison, WI 53706 USA

[b]Department of Crop Sciences, University of Illinois, 1101 W. Peabody Drive, Urbana, IL, 61801 USA

[1]To whom correspondence may be addressed: Email: afbent@wisc.edu.

Keywords: Plant disease resistance, α-SNAP, NSF, soybean cyst nematode, *Rhg1*

2.1 Abstract

NSF and α-SNAP (N-ethylmaleimide Sensitive Factor; α-Soluble NSF Attachment Protein) are essential eukaryotic housekeeping proteins that cooperatively function to sustain vesicular trafficking. The Resistance to *Heterodera glycines* 1 (*Rhg1*) locus of soybean (*Glycine max*) confers resistance to soybean cyst nematode, a highly damaging soybean pest. *Rhg1* loci encode repeat copies of atypical α-SNAP proteins that are defective in promoting NSF function and are cytotoxic in certain contexts. Here, we discovered an unusual *NSF* allele (*Rhg1*-associated NSF on chromosome 07; NSF_{RAN07}) in $Rhg1^{+}$ germplasm. NSF_{RAN07} protein modeling to mammalian NSF/α-SNAP complex structures indicated that at least three of the five NSF_{RAN07} polymorphisms reside adjacent to the α-SNAP binding interface. NSF_{RAN07} exhibited stronger *in vitro* binding with *Rhg1* resistance-type α-SNAPs. NSF_{RAN07} co-expression *in planta* was more protective against *Rhg1* α-SNAP cytotoxicity, relative to WT NSF_{Ch07}. Investigation of a previously reported segregation distortion between chromosome 18 *Rhg1* and a chromosome 07 interval now known to contain the *Glyma.07G195900* NSF gene revealed 100% co-inheritance of the NSF_{RAN07} allele with disease resistance *Rhg1* alleles, across 855 soybean accessions and in all examined $Rhg1^{+}$ progeny from biparental crosses. Additionally, we show that some *Rhg1*-mediated resistance is associated with depletion of WT α-SNAP abundance via selective loss of WT α-SNAP loci. Hence atypical co-evolution of the soybean SNARE-recycling machinery has balanced the acquisition of an otherwise disruptive housekeeping protein, enabling a valuable disease resistance trait. Our findings further indicate that successful engineering of *Rhg1*-related resistance in plants will require a compatible NSF partner for the resistance-conferring α-SNAP.

2.2 Significance Statement

NSF and α-SNAP proteins are key components of vesicle trafficking systems and are conserved across eukaryotes. This study shows that these two essential housekeeping proteins have co-evolved toward atypical forms in soybean to confer resistance to a highly damaging nematode pathogen while balancing plant fitness. We report discovery of a naturally occurring NSF variant carrying unusual polymorphisms that enhance interaction with and assuage the cytotoxicity of the *Rhg1* resistance-associated α-SNAPs. Pathogen selection pressure has apparently driven this re-wiring of multiple components of the conserved SNARE recycling machinery. Useful introduction of the agriculturally valuable *Rhg1* resistance source into other plants is likely to require a co-functional NSF protein partner.

2.3 Introduction

Cyst nematodes infest the roots of many valuable crops and establish elaborate feeding structures (Kyndt et al., 2013). Soybean cyst nematode (*Heterodera glycines*; SCN) is a highly damaging soybean pest and causes annual U.S. yield losses of over $1 billion USD (Niblack et al., 2006; Jones et al., 2013; Mitchum, 2016; Allen et al., 2017). SCN parasitizes host roots by secreting a complex arsenal of effector molecules that reprogram host root cells and trigger fusion with adjacent host cells, forming a large unicellular feeding site termed a syncytium (Gheysen and Mithcum, 2011; Hewezi and Baum, 2013; Mitchem et al., 2013). The soybean *Rhg1* (Resistance to *Heterodera glycines* 1) locus is very widely used by soybean growers to restrict SCN feeding site formation, thereby reducing yield loss (Concibido et al., 2004; Mitchum, 2016). The genes at *Rhg1* do not encode proteins normally associated with disease resistance (Dodds and Rathjen, 2010; Cook et al., 2012; Lee et al., 2015; Mitchum, 2016). Instead, resistance is mediated by copy number variation of multiple genes at the *Rhg1* locus, one of which encodes an α-SNAP with unusual C-terminal polymorphisms (Cook et al., 2012; Cook et al., 2014; Lee et al., 2015) .

α-SNAP (alpha-Soluble NSF-Attachment Protein; Sec17 in yeast) is a functionally conserved eukaryotic housekeeping protein that works in concert with NSF (N-ethylmaleimide Sensitive Factor; Sec18 in yeast). α-SNAP and NSF promote cellular vesicular trafficking by mediating the disassembly and reuse of SNARE protein complexes (Soluble NSF Attachment protein Receptors) that form when t-SNARE and v-SNARE proteins associate during vesicle docking and fusion (jahn and Scheller, 2006; Baker and Hugson, 2016; Zhao and Brunger, 2016). We recently discovered that the soybean resistance-associated α-SNAPs encoded by *Rhg1* are

unusual α-SNAP proteins that bind less well to wild-type (WT) NSF and, when expressed in *Nicotiana benthamiana*, disrupt vesicle trafficking and eventually cause cell death (Bayless et al., 2017). The relative abundance of *Rhg1*-encoded defective α-SNAP variants increases substantially within developing host syncytial cells, apparently disrupting syncytium viability and thereby restricting nematode growth and reproduction (Bayless et al., 2016). SCN-resistant soybeans carry wild-type α-SNAP genes at other loci that can functionally complement the *Rhg1* resistance-type α-SNAPs in a dosage-dependent manner (Bayless et al., 2016). However, the capacity of soybean varieties to yield well despite expression of cytotoxic *Rhg1* resistance-type α-SNAPs throughout the plant is not fully explained.

The complex *Rhg1* locus on soybean chromosome 18 is a tandemly repeated block of four genes: *Glyma.18G022400*, *Glyma.18G022500*, *Glyma.18G022600* and *Glyma.18G022700*. SCN-susceptible soybeans carry only a single copy of the above four genes, including a *Glyma.18G022500* α-SNAP gene whose product matches the wild-type (WT) α-SNAP consensus and maintains normal NSF interactions (Cook et al., 2012; Cook et al., 2014; Bayless et al., 2016). Resistance-conferring *Rhg1* loci group into two structural classes based on the type of α-SNAP polymorphisms they encode, which also correlates with the copy-number of *Rhg1* repeats that are present (Cook et al., 2014; Lee et al., 2015) (Table S1). *Rhg1*$_{HC}$ (high copy) loci carry four or more and frequently nine or ten *Rhg1* repeats, and *Rhg1*$_{LC}$ (low-copy) loci carry three or fewer *Rhg1* repeats. *Rhg1*$_{HC}$ (also known as *rhg1-b*) and *Rhg1*$_{LC}$ (also known as *rhg1-a*) encode distinct α-SNAP variants that are impaired in normal α-SNAP-NSF interactions (Bayless et al., 2016)(Fig. 1*A*). All *Rhg1*$_{HC}$ loci examined to date also carry a single *Rhg1* repeat that encodes a WT α-SNAP adjacent to multiple repeats that encode resistance-type α-SNAPs, while *Rhg1*$_{LC}$ loci

encode only resistance-type α-SNAPs and no WT α-SNAP (Cook et al., 2012; Cook et al., 2014; Lee et al., 2015)(Fig. 1*A*). Plants carrying *Rhg1*$_{HC}$ or *Rhg1*$_{LC}$ loci exhibit elevated transcript abundance for the repeat genes that correlates approximately with copy number, including for the *Rhg1* α-SNAP gene (Cook et al., 2012; Cook et al., 2014). Collectively, the above findings suggest that modulation of vesicle trafficking and cell health at the SCN feeding site is at least one core mechanism of *Rhg1*-mediated SCN resistance.

Two other genes within the *Rhg1* repeat were reported by Cook *et al.* to contribute to *Rhg1*$_{HC}$-mediated SCN resistance (Cook et al., 2012). *Glyma.18G022400* encodes an amino acid permease-like protein and *Glyma.18G022700* encodes a wound-inducible protein lacking annotated domains or predicted functions; their molecular function in SCN resistance remains unknown. Liu *et al.* recently provided evidence that the *Rhg1*$_{LC}$ α-SNAP may function differently than the *Rhg1*$_{HC}$ α-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). The role of α-SNAP and NSF as dedicated SNARE-recycling chaperones has been studied extensively (Jahn and Scheller, 2006; Wickner and Schekman, 2008; Wickner, 2010; Zhao et al., 2015; Zick et al., 2015) . NSF is an AAA^{+} family protein (ATPases Associated with various cellular Activities) with 3 domains: the N-domain that binds and interacts with the C-terminus of the α-SNAP co-chaperone, the D1 ATPase domain that couples ATP hydrolysis to SNARE complex remodeling, and the D2 ATPase domain that

mediates NSF hexamerization (Whiteheart et al., 2001; Hanson and Whiteheart, 2005; Zhao et al., 2010). α-SNAP proteins are required by NSF to co-chaperone SNARE remodeling. α-SNAP serves both as an adaptor for NSF binding to SNARE complexes and as a stimulator of the NSF D1 domain ATPase activity that powers SNARE remodeling/recycling (Zhao and Brunger, 2016). Beyond disassembling SNARE complexes, additional roles of α-SNAP and NSF have been reported, including 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 Mateeva, 2004; Zhao et al., 2007; Naydenov et al., 2012; Miao et al., 2013; Zick et al., 2015) . The structure and function of α-SNAP, NSF and SNARE proteins has been elucidated in substantial detail, including cryo-EM structures for 20S complexes that consist of a four-helix SNARE bundle, four α-SNAPs and six NSFs subunits in various conformational states (Zhao et al., 2015; Zhao and Brunger, 2016).

Although most animal genomes carry a single NSF and a single α-SNAP gene, polyploidization and other events have caused most plant genomes to encode multiple NSF and α-SNAP genes (Littleton et al., 2001). The reference Williams 82 soybean genome encodes seven Soluble NSF Attachment Protein (SNAP) family members: five putative α-SNAPs and 2 putative γ-SNAPs. Soybean also encodes two unlinked NSF genes, *Glyma.07G195900* and *Glyma.13G180100*. 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 (including SNAREs with potentially confusing names such as synaptosomal-associated protein 25 (SNAP-25) and soluble N-ethylmaleimide-sensitive factor adaptor protein 33 (SNAP33)), there are very few published studies of plant NSF and SNAP proteins (Bassham and

Raikhel, 1999; Bachem et al., 2000; Rancour et al., 2002; Cook et al., 2012; Matsye et al., 2012; Cook et al., 2014; Bayless et al., 2016; Liu et al., 2017) . However, close analysis of recombinant-inbred lines has recently shown that a gene at or linked to the soybean Chromosome 11 locus encoding an α-SNAP makes a minor contribution to SCN resistance in the Peking (*Rhg1*$_{LC}$ + *Rhg4*) genetic background (Lakhssassi et al., 2017). Other previous work (Matsye et al., 2012) had identified an allele encoding a splice-variant α-SNAP in this genetic background, although that work misidentified it as an allele of the Chromosome 18 *Rhg1* locus despite it now being known to be a Chromosome 11 α-SNAP allele (Cook et al., 2014; Lakhssassi et al., 2017).

In the present study, we demonstrate that evolution/selection of both *Rhg1*$_{LC}$ and the Chromosome 11 α-SNAP gene *Glyma.11G234500* has had major impacts on the relative abundance of WT α-SNAP proteins in the *Rhg1*$_{LC}$ genetic background. We also examined soybean NSF proteins. We discovered an unusual NSF protein in *Rhg1*-containing lines that is unlike that encoded in the soybean Williams 82 reference genome or any publicly available plant reference genomes. We found that this variant *NSF*$_{RAN07}$ (*Rhg1*-associated NSF on Chromosome 07; NSF$_{RAN07}$) protein contains unique N-domain polymorphisms that mitigate the cytotoxicity and poor NSF binding activity of the SCN resistance-conferring *Rhg1* α-SNAPs. We then noted that the genetic region containing this NSF and neighboring genes has been identified in previous SCN resistance mapping studies, including a 1995 study by Webb showing strong co-segregation with resistance-conferring *Rhg1* alleles (Webb et al., 1995; Kopisch-Obuch and Diers. 2006). More recently a high-resolution 80kb candidate gene interval was identified (Vuong et al., 2015) but this segregation distortion at the Chromosome 07 locus had

remained unexplained. We therefore investigated soybean germplasm genotype data and recombinant inbred lines from *Rhg1*$^{+}$ x *rhg1*$^{-}$ parental crosses. We discovered strict co-inheritance of NSF_{RAN07} alleles in plants homozygous for resistance-associated *Rhg1* haplotypes, demonstrating the functional necessity of NSF_{RAN07} for viable occurrence of SCN resistance-conferring *Rhg1*.

2.4 Results

Wild-type α-SNAP proteins are much less abundant while NSF is more abundant in *Rhg1*$_{LowCopy}$ soybeans

We previously reported that the PI 88788-type high copy *Rhg1* (*Rhg1*$_{HC}$) locus in soybean line "Fayette" drives a localized increase of resistance-type α-SNAP$_{Rhg1}$HC protein to disrupt the developing SCN-induced syncytium (Bayless et al., 2016). We also observed that endogenous NSF levels increased when resistance-associated *Rhg1* α-SNAP proteins were overexpressed in *N. benthamiana* (Bayless et al., 2016). However, for lines carrying low-copy type *Rhg1* (*Rhg1*$_{LC}$, "Peking-type"), the cellular balance of WT α-SNAP to α-SNAP$_{Rhg1}$LC or NSF proteins was unknown. To investigate the relative abundances of WT and resistance-associated α-SNAPs, we used previously described anti-α-SNAP antibodies and performed immunoblots on the *Rhg1*$_{HC}$ and *Rhg1*$_{LC}$ soybean varieties commonly used to phenotype SCN-resistance (the HG Type Test varieties; see Table S1) (Niblack et al., 2002; Bayless et al., 2016). We also examined the abundance of the α-SNAP co-chaperone NSF in these samples, using an antibody raised to a conserved NSF region (Bayless et al., 2016). Fig. 1*A* presents a schematic of the various *Rhg1* haplotypes as well as the C-terminal polymorphisms of *Rhg1* α-SNAPs encoded by the *Rhg1*

repeat types. As shown in Figure 1B, immunoblots from root tissue indicated that WT α-SNAP protein levels in all tested *Rhg1*$_{LC}$ lines (PI 548402/Peking, PI 90763, PI 437654, PI 89772) are dramatically reduced compared with the *Rhg1*$_{HC}$ lines (PI 88788, PI 209332, PI 548316). As mentioned above, the Williams 82 (Wm82) soybean genome encodes five putative α-SNAPs, and the anti-WT-α-SNAP antibody was raised against the conserved C-terminus shared by all of those predicted WT α-SNAP gene products but not the resistance-associated *Rhg1* α-SNAPs (17). In addition, one *Rhg1* repeat in *Rhg1*$_{HC}$ haplotypes encodes a WT *Glyma.18G022500* α-SNAP protein and all other repeats encode a resistance-type *Rhg1* α-SNAP protein, while the *Rhg1*$_{LC}$ repeats encode only resistance-type α-SNAP$_{Rhg1}$LC proteins (Fig. 1*A*) (Cook et al., 2014; Lee et al., 2015). The results of Fig. 1B did not match initial predictions; the tested *Rhg1*$_{LC}$ soybean lines exhibit very low WT α-SNAP protein levels despite the presence of multiple α-SNAP genes at other loci.

We further discovered that total NSF protein abundance in the *Rhg1*$_{LC}$ lines is increased compared with the *Rhg1*$_{HC}$ lines PI 88788 and PI 209332 (Fig. 1*B*, Fig. S1*A*). These differences in NSF abundance, across two independent experiments, were quantified using densitometry (Fig. 1*C*).

We then explored if WT α-SNAP protein abundance is similarly reduced in a more recent agriculturally utilized *Rhg1*$_{LC}$ soybean variety, "Forrest." Immunoblots on both total leaf or root proteins from Williams 82 (*Rhg1* single copy), Forrest (*Rhg1*$_{LC}$) and Fayette (*Rhg1*$_{HC}$) again revealed sharp decreases in total WT α-SNAP abundance in the *Rhg1*$_{LC}$ source (Fig. 1*D*). Altogether, diminished WT α-SNAP protein levels were observed to be a shared trait of *Rhg1*$_{LC}$ but not *Rhg1*$_{HC}$ soybean varieties. In at least two previously studied *Rhg1*$_{LC}$ varieties, as well as

PI 548316, the Chromosome 11 α-SNAP allele (*Glyma.11G234500*) carries a SNP at an intronic splice donor site (Table S1, Fig. 1*F*), leading to intron-retention and early translational termination, presumably truncating the protein (Cook et al., 2014; Matsye et al., 2012; Lakhssassi et al., 2017). Hence a likely hypothesis for this strikingly low abundance is the absence of a WT-α-SNAP-encoding allele at *Rhg1*$_{LC}$, low or no product from the α-SNAP$_{Ch11}$ allele whose transcript retains a translation-terminating intron, and a relatively minor contribution of protein from the other three putative α-SNAP-encoding loci.

Contributions to WT α-SNAP abundance were investigated further. First, we examined overall WT α-SNAP protein abundance when a locus encoding α-SNAP$_{Rhg1}$WT is ectopically placed into *Rhg1*$_{LC}$ soybean lines. We cloned from Wm82 the genomic chromosome 18 (Ch18) *Glyma.18G022500* α-SNAP$_{Rhg1}$WT locus with its native promoter and terminator sequences, generated transgenic Forrest (*Rhg1*$_{LC}$) roots carrying this native α-SNAP$_{Rhg1}$WT locus, and assessed total WT α-SNAP protein levels using immunoblots (Fig. 1*E*). Transgenic addition of the Wm82 α-SNAP$_{Rhg1}$WT locus increased total WT α-SNAP protein expression in Forrest to levels similar to Wm82 empty vector controls (Fig. 1*E*). This result indicates that if an appropriate gene is present, normal WT α-SNAP protein levels can develop in the *Rhg1*$_{LC}$ genetic background.

Next, we examined α-SNAP protein production from the chromosome 11 (Ch11) WT locus from Wm82 vs. the Ch11 intron-retention allele (*α-SNAP$_{Ch11}$-IR*) that is present in many soybean lines that carry *Rhg1$_{LC}$* on Ch18. The transcript from the intron-retention allele encodes a premature stop codon (Cook et al., 2014; Matsye et al., 2012; Lakhssassi et al., 2017) (Fig. 1*F*), but the abundance/stability of this putative α-SNAP protein was not known. As such,

we cloned open reading frames of both the wild-type α-SNAP$_{Ch11}$ and the intron-retention (α-SNAP$_{Ch11}$-IR) alleles, added an N-terminal HA tag, and examined transient protein expression in *N. benthamiana*. We observed that the HA-α-SNAP$_{Ch11}$ WT protein, but not the truncated HA-α-SNAP$_{Ch11}$-IR protein, was readily detectable (Fig. 1*G*). Finally, as for the Ch18 locus tested in Fig. 1*E*, we cloned the Ch11 genomic WT locus of *Glyma.11G234500* (α-SNAP$_{Ch11}$) from Wm82 with native promoter and terminator and noted that presence of this native locus in transgenic roots of Forrest elevated total WT α-SNAP protein expression compared to empty vector controls (Fig. S1*C*). Together, the findings of Fig. 1 and Fig. S1 implicate the Ch18 and Ch11 WT α-SNAP loci as the major sources of total WT α-SNAP proteins in soybean and indicate that their combined absence from the examined *Rhg1*$_{LC}$ varieties is responsible for the low levels of WT α-SNAP observed in Fig. 1*B* and 1*D*. The low abundance of WT α-SNAPs in lines carrying *Rhg1*$_{LC}$ may improve SCN resistance but may also incur costs with respect to plant health and yield if other compensatory mechanisms for tolerance of *Rhg1*$_{LC}$ are not also present.

A unique *NSF$_{Ch07}$* allele (*NSF$_{RAN07}$*) is present in commonly used *Rhg1*-containing accessions

NSF and α-SNAP are essential eukaryotic housekeeping proteins and null mutations in either partner is lethal in animals, which typically encode only single copies of NSF or α-SNAP (Littelton et al., 2001; Sanyal and Krishnan, 2001; Horsnell et al., 2002; Chae et al., 2004) . Because *Rhg1*-resistance type α-SNAPs (α-SNAP$_{Rhg1}$LC or α-SNAP$_{Rhg1}$HC) exhibit compromised binding to wild-type NSFs and are toxic at high doses in *N. benthamiana* , (Bayless et al., 2017) it was unclear how *Rhg1*$_{LC}$ lines are viable given the diminished WT α-SNAP levels observed in Figure 1. Since soybean is an ancestrally polyploid organism encoding multiple α-SNAP and NSF

loci, we searched for alterations in the other α-SNAP or NSF loci by examining our previously generated whole genome sequence (WGS) data from multiple *Rhg1*-containing varieties (Cook et al., 2014). For all five putative α-SNAP loci from *Rhg1*$_{LC}$ varieties, we detected no obvious polymorphisms other than the previously mentioned *Glyma.11G234500* intron-retention allele (Table S1, Table S2) (Cook et al., 2014; Lakhssassi et al., 2017).

Intriguingly, a novel NSF allele was present at *Glyma.07G195900* (NSF$_{Ch07}$) among all six of the *Rhg1*$_{LC}$ and *Rhg1*$_{HC}$ lines examined, encoding five N-domain amino acid polymorphisms (R_4Q, $N_{21}Y$, $S_{25}N$, $\wedge_{116}F$, $M_{181}I$; ^ = insertion) (Fig. 2*A*, S2*A*, Table S1). Using cDNA from Forrest (*Rhg1*$_{LC}$), we cloned and sequenced this unique NSF$_{Ch07}$ transcript and confirmed the five N-domain polymorphisms. Additionally, we designed two different PCR primer pairs at the encoded NSF polymorphisms and verified the presence of this unique NSF $_{Ch07}$ allele, and the absence of the wild-type NSF$_{Ch07}$ allele, in all *Rhg1* HG test lines (Fig. S2*B*). Furthermore, using WGS data from the SoyNAM (Nested Association Mapping) project (Song et al., 2017), we determined that this unique NSF$_{Ch07}$ allele was also present in every *Rhg1*-containing NAM parent, while SCN-susceptible NAM parents carried the WT NSF$_{Ch07}$ allele (Table S2). We therefore named the protein from this *Rhg1*-associated allele of *Glyma.07G195900* "NSF$_{RAN07}$" for "*Rhg1*-associated NSF on chromosome 07." In addition to *NSF$_{RAN07}$*, an allele of the chromosome 13 *Glyma.13G180100* gene encoding an NSF$_{Ch13}$ $V_{555}I$ protein was found in some varieties, including SCN-susceptible soybeans, but it was not present in every *Rhg1*$_{LC}$ or *Rhg1*$_{HC}$ line (Table S2). Normalized RNA seq. reads from Williams 82 indicates that both *Glyma.07G195900* and *Glyma.13G180100* are expressed similarly across examined plant tissues

(Fig. S2*C*) (Severin et al., 2010). Figure S2*A* provides the complete NSF_{RAN07} amino acid alignment to NSF_{Ch07} from the Williams 82 genome.

The NSF_{RAN07} and *Rhg1* α-SNAP polymorphisms lie at the NSF/α-SNAP binding interface

Multiple α-SNAP proteins bind to a SNARE bundle and recruit NSF proteins (most typically 4 SNAREs, 4 α-SNAPs and 6 NSF proteins), to form a "20S supercomplex" and stimulate SNARE complex disassembly (Zhao et al., 2015). The NSF/α-SNAP interface consists of complementary electrostatic patches located at the NSF N-domain and α-SNAP C-terminus (Zhao et al., 2015; Zhao and Brunger, 2016) The *Rhg1* polymorphisms of both α-$SNAP_{Rhg1}$HC and α-$SNAP_{Rhg1}$LC are located at conserved C-terminal residues shown in other α-SNAPs to bind and stimulate NSF (Barnard et al., 1996; Cook et al., 2014; Bayless et al., 2016). These binding patches are conserved in yeast, animals and plants, and inter-kingdom interactions between α-SNAP and NSF have been reported between mammals and yeast and plants, including soybean WT α-SNAP and Chinese hamster NSF (NSF_{CHO}) (Griff et al., 1992; Bassham and Raikhel, 1999; Rancour et al., 2002; Bayless et al., 2016). We performed homology modeling of NSF_{RAN07} to the NSF_{CHO} cryo-EM structure (Zhao et al., 2015) (PDB 3j97.1) that placed three of the NSF_{RAN07} polymorphisms, $N_{21}Y$, $S_{25}N$ and the $\wedge_{116}F$ insertion, adjacent to the NSF_{CHO} α-SNAP-binding residues R_{10} and $RK_{104\text{-}105}$ (Fig. 2*B*, S3*A*). NSF_{RAN07} polymorphism R_4Q was outside of the model and the final NSF_{RAN07} polymorphism $M_{181}I$ was not located near the α-SNAP binding patches. Further homology modeling was conducted using the mammalian 20S cryo-EM structure (PDB 3j97). In Fig. 2*C* and Fig. S*4A*, *B* the complementary NSF and α-SNAP binding residues, and the NSF_{RAN07} and *Rhg1* α-SNAP polymorphisms, are colored. These results suggest that upon α-

SNAP binding, NSF_{RAN07} $N_{21}Y$, $S_{25}N$ and $^{}_{116}F$ are close to the WT α-SNAP amino acid residues that are polymorphic in α-$SNAP_{Rhg1}$HC and α-$SNAP_{Rhg1}$LC. In separate bioinformatics work we examined the NSF N-domain consensus in plants and determined that residues corresponding to N_{21} and F_{115} of WT soybean NSF are present in a majority of plant species, while neither the $N_{21}Y$ nor the $^{}_{116}F$ insertion of NSF_{RAN07} were detected in any available plant reference genome sequences (Fig. S3*B*). Altogether, this modeling suggested that NSF_{RAN07} carries rare alterations at the α-SNAP binding interface that potentially influence interactions with the unusual C-termini of *Rhg1* resistance-type α-SNAPs.

NSF_{RAN07} polymorphisms enhance binding with *Rhg1* resistance-type α-SNAPs

In light of the above, NSF_{RAN07} binding with *Rhg1* resistance-type α-SNAPs and α-$SNAP_{Rhg1}$WT was investigated. As in (Barnard et al, 1997) and (Bayless et al., 2017), we produced recombinant NSF_{RAN07}, NSF_{Ch07} and *Rhg1* α-SNAP proteins and performed *in vitro* binding assays. NSF_{RAN07} and NSF_{Ch07} binding was quantified using densitometry across three independent experiments (Fig. 2*E*). As previously reported (Bayless et al., 2017), diminished NSF_{Ch07} binding was observed for α-$SNAP_{Rhg1}$HC and α-$SNAP_{Rhg1}$LC, compared to α-$SNAP_{Rhg1}$WT (Fig. 2*D*). α-$SNAP_{Rhg1}$HC or α-$SNAP_{Rhg1}$LC binding of NSF_{RAN07}, on the other hand, was more similar to α-$SNAP_{Rhg1}$WT binding of NSF_{RAN07} and was increased ~30% relative to the binding of NSF_{Ch07} (Fig. 2*D*, *E*).

To investigate the contribution of the α-SNAP C-terminus to NSF_{RAN07} binding, we tested NSF_{RAN07} binding to an otherwise WT α-SNAP that lacked the final 10 C-terminal residues (α-$SNAP_{Rhg1}WT_{1\text{-}279}$). Similar to the "no α-SNAP" binding controls, essentially no binding of either

NSF_{Ch07} or NSF_{RAN07} with α-$SNAP_{Rhg1}WT_{1-279}$ was observed (Fig. S4*D*). To more specifically investigate the NSF binding contribution of just the C-terminal residues polymorphic in α-$SNAP_{Rhg1}$LC (see Fig. 1*A*), we mutagenized α-$SNAP_{Rhg1}$LC from $286_{YEVI}289$ to $286_{AAAA}289$. Binding of either NSF_{Ch07} or NSF_{RAN07} to α-$SNAP_{Rhg1}$LC $286_{AAAA}289$ was similar to "no α-SNAP" controls (Fig. S4*D*, *E*). Hence NSF_{RAN07} binding is sensitive to the α-SNAP C-terminal residues that are polymorphic in the *Rhg1* resistance-type α-SNAPs.

We then examined if two of the key NSF_{RAN07} polymorphisms (Y_{21} and F_{116}) that model near predicted α-SNAP binding patches influence binding to *Rhg1* α-SNAPs. We restored these two residues back to the identities in WT NSF_{Ch07}, while retaining the other three NSF_{RAN07} polymorphisms (Q_4, N_{25}, I_{181}). Performing *in vitro* binding assays as above, we observed a reduced ability of NSF_{RAN07} $Y_{21}N$ F_{116}^, as compared to unaltered NSF_{RAN07}, to bind resistance-type α-SNAPs (Fig. 2*F*, S4*E*). Mutating these two positions to alanine in an otherwise WT NSF_{Ch07} (NSF_{Ch07} $N_{21}A$ $F_{115}A$) did not restrict binding with WT α-SNAP, and binding of this NSF_{Ch07} $N_{21}A$ $F_{115}A$ with either α-$SNAP_{Rhg1}$HC or α-$SNAP_{Rhg1}$LC was still impaired (Fig. 2*F*, S4*D*). Combined, these *in vitro* binding results suggest that NSF_{RAN07} not only maintains normal binding with WT α-SNAPs, but can also accommodate the unusual C-terminal polymorphisms of the *Rhg1* resistance-type α-SNAPs.

The NSF_{RAN07} polymorphisms guard against the cell death induced by *Rhg1*-resistance-type α-SNAPs

We previously observed that transient expression of either α-$SNAP_{Rhg1}$HC or α-$SNAP_{Rhg1}$LC in *N. benthamiana* leaves, via *Agrobacterium* infiltration, is cytotoxic and elicits a

hyperaccumulation of the endogenous NSF protein (Bayless et al., 2016). Co-expression of a WT-α-SNAP with the *Rhg1* resistance-type α-SNAP diminishes this toxicity in a dose-dependent manner, and also relieves negative impacts on sec-GFP secretion (Bayless et al., 2016). The penultimate amino acid (conserved leucine) of α-SNAP, which has been implicated in stimulation of NSF ATPase, is needed for rescue of this *N. benthamiana* cytotoxicity (Barnard et al., 1997; Zick et al., 2015; Bayless et al., 2016). We subsequently conducted site-directed mutagenesis experiments which provided further evidence that the *N. benthamiana* assay closely correlates with known α-SNAP/NSF behaviors. In a first set of replicated studies, the toxicity of *Rhg1* α-SNAP expression, and the capacity of co-expressed WT α-SNAP to protect against *Rhg1* α-SNAP toxicity were both observed to be dependent on intact SNARE-binding sites within the respective α-SNAPs (Fig. S5).

We then examined if, like WT α-SNAP, co-expression of soybean NSF might alleviate the toxicity of *Rhg1* resistance-type α-SNAPs in *N. benthamiana*. Similar to (Bayless et al., 2017), mixed *Agrobacterium* inocula were used, with ratios varying from 1:4 (1-part NSF-expressing strain *t*o 4-parts α-SNAP$_{Rhg1}$LC-expressing strain) all the way down to 1:19. NSF co-expression strongly reduced *Rhg1* α-SNAP cytotoxicity (Fig. 3 and Fig. S6*A*). No macroscopic phenotypes indicative of stress were observed upon expressing NSF$_{RAN07}$ or NSF$_{Ch07}$ alone (Fig. S6*A*). Titration of the dose-response for NSF-expressing *Agrobacterium* strains identified a range of effective strain ratios (c.f., Fig. 3*B*). We observed that co-expressing soybean NSF$_{Ch07}$, NSF$_{Ch13}$ or NSF$_{RAN07}$ reduced cell death caused by α-SNAP$_{Rhg1}$LC compared to empty vector controls (Fig. 3*A, B*). However, NSF$_{RAN07}$ co-expression consistently conferred greater protection than either NSF$_{Ch07}$ or NSF$_{Ch13}$ (Fig. 3*A, B*). Across multiple independent sets of leaves tested at a variety of

ratios, we observed that leaf patches co-infiltrated with NSF_{RAN07} exhibited less cell death and/or slower death. Protection against α-$SNAP_{Rhg1}$HC-induced cell death with NSF_{RAN07} vs. NSF_{Ch07} produced similar results (Fig. S6*B*).

As noted above, we have consistently observed elevated abundance of the endogenous *N. benthamiana* NSF ($NSF_{N.benth}$) upon expression of *Rhg1* resistance-type α-SNAPs, yet this does not prevent cell death (Bayless et al., 2016) (Fig. 1). However, it was unclear if immediate co-expression of $NSF_{N.benth}$ (81% identity to soybean NSF_{Ch07}, see Fig. S7 for alignment) might lessen the cytotoxicity. Therefore, we agroinfiltrated mixed cultures expressing $NSF_{N.benth}$ and α-$SNAP_{Rhg1}$LC, as well as empty-vector, NSF_{Ch13} and NSF_{RAN07} as controls. As in Fig. 3*A*, NSF_{Ch13} coexpression gave partial protection while NSF_{RAN07} coexpression gave strong protection (Fig. 3*C*). $NSF_{N.benth}$ coexpression, on the other hand, was similar to empty vector controls and did not guard against α-$SNAP_{Rhg1}$LC- induced cell death (Fig. 3C). Because no obvious cell-death rescue from co-expressing $NSF_{N.benth}$ was apparent, we also examined $NSF_{N.benth}$ physical binding with *Rhg1* resistance-type α-SNAPs, using recombinant $NSF_{N.benth}$ protein. $NSF_{N.benth}$ readily bound α-$SNAP_{Rhg1}$WT, but binding to either *Rhg1* resistance-type α-SNAP was much lower, only slightly over negative controls (Fig. 3*D*). These experiments suggest that $NSF_{N.benth}$ exhibits little or no functional interaction with SCN resistance-associated soybean *Rhg1* α-SNAPs, which likely accounts for the high toxicity of *Rhg1* α-SNAPs in *N. benthamiana*.

We then used the *N. benthamiana* assay to examine NSF_{RAN07} function predictions. One set of experiments tested if cell-death caused by α-$SNAP_{Rhg1}LC_{1\text{-}279}$, which lacks the final 10 C-terminal residues, could be rescued by NSF_{RAN07} or NSF_{Ch07}. Neither NSF_{RAN07} nor NSF_{Ch07} prevented the cell death caused by α-$SNAP_{Rhg1}LC_{1\text{-}279}$, despite guarding against cell death in the

positive control treatments involving full length α-SNAP$_{Rhg1}$LC (Fig. S8*A*). Likewise, we tested if cell death caused by α-SNAP$_{Rhg1}$LC 286$_{AAAA}$289 - which also did not exhibit *in vitro* binding of NSF – could be rescued by either NSF$_{RAN07}$ or NSF$_{Ch07}$. α-SNAP$_{Rhg1}$LC 286$_{AAAA}$289, like α-SNAP$_{Rhg1}$LC, elicited increased expression of the endogenous *N. benthamiana* NSF (Fig. S8*B*,*C*). Yet compared with α-SNAP$_{Rhg1}$LC, which does bind the tested soybean NSF to some extent, we observed that α-SNAP$_{Rhg1}$LC 286$_{AAAA}$289-induced cell death was not strongly protected by NSF$_{RAN07}$ or NSF$_{Ch07}$ co-expression (Fig. S8*B*). These experiments provide further evidence that C-terminally mutagenized α-SNAPs also disrupt *N. benthamiana* 20S complexes and, that NSF rescue of the cell death induced by toxic a-SNAPs requires an intact α-SNAP C-terminus to mediate successful α-SNAP-NSF interaction.

Turning to the NSFs mutagenized at the inferred α-SNAP binding interface, α-SNAP$_{Rhg1}$LC cell death rescue via co-expression of mutated NSF$_{Ch07}$ or NSF$_{RAN07}$ (NSF$_{Ch07}$ N$_{21}$A F$_{115}$A or NSF$_{RAN07}$ Y$_{21}$A F$_{116}$^) was not as robust as rescue by the normal NSF$_{Ch07}$ or NSF$_{RAN07}$ (Fig. S*8F, G*). Anti-NSF immunoblots confirmed the expression of NSF$_{Ch07}$, NSF$_{RAN07}$ and their respective mutants (Fig. S8*E*). This supports the contribution of the mutated NSF residues to optimal NSF/α-SNAP interaction.

Finally, we made and used an α-SNAP$_{Rhg1}$LC I$_{289}$A to examine how the penultimate α-SNAP residue, which has been shown in other α-SNAPs to help stimulate NSF-ATPase, affected rescue by NSF$_{RAN07}$ or NSF$_{Ch07}$ (Barnard et al., 1996; Zick et al., 2015). Protection against α-SNAP$_{Rhg1}$LC I$_{289}$A was evident but was much less than that observed for α-SNAP$_{Rhg1}$LC (Fig. S8*D*), suggesting that although NSF$_{RAN07}$ may bind *Rhg1* resistance type α-SNAPs more effectively, ATPase-stimulation is likely an additional factor in relieving cytotoxicity. Overall, the findings of

Figure 3 extend the Figure 2 finding that NSF_{RAN07} binds *Rhg1* α-SNAPs, demonstrating *in vivo* that the NSF_{RAN07} polymorphisms more effectively guard against the disruptive effects of the polymorphic *Rhg1* α-SNAPs, and demonstrating that, among site-directed mutants, the extent of this *in planta* protection correlates with observed *in vitro* α-SNAP-NSF binding differences.

One hundred percent of predicted *Rhg1*$^+$ *Glycine max* accessions in the USDA soybean collection contain the NSF_{RAN07} R_4Q amino acid polymorphism

NSF_{RAN07} was present in all *Rhg1*-containing HG type and NAM lines (Table S1, Table S2), but we sought to test if this *Rhg1*/NSF_{RAN07} association was universal rather than "frequent". 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). The data are available in a searchable SNP database at Soybase (Soybase.org/snps/) (Grant et al., 2010; Song et al., 2013; Song et al., 2015). Using this Soybase SNP browser, we found that a C/T SNP (ss715597431) causes the NSF_{RAN07} R_4Q polymorphism. Analyzing all 19,645 USDA *Glycine max* accessions for ss715597431, we estimated the NSF_{RAN07} allele frequency in the USDA collection at 11.0% (2,165 NSF_{RAN07}^+/ NSF_{RAN07}^+, 33 NSF_{RAN07}^+/ NSF_{RAN07}^-) (Fig. 4*A*). Q_4 was not found in the predicted NSF protein sequences of any plant species available for query at Phytozome.org (Goodstein et al., 2012) (Fig. S9).

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 $Rhg1_{HC}$, $Rhg1_{LC}$ and single copy (SCN-susceptible) haplotypes, and estimated that 705 $Rhg1_{LC}$ and 150 $Rhg1_{HC}$ accessions were present in the

USDA *Glycine max* collection (Lee et al., 2015). Among these 855 *Rhg1*-signature *Glycine max* accessions, we determined a 100% incidence of the ss715597431 NSF_{RAN07} signature (Fig. 4*B*).

To better define the *Rhg1* co-segregating locus within the Ch07 interval, we examined amino acid changes within candidate loci adjacent to NSF_{RAN07} from *Rhg1*-carrying HG and NAM lines, between markers ss715597415 and ss715597431. We observed that the NSF_{RAN07} SNPs, especially those causing the five polymorphisms in the N-domain, 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 S3). The predicted amino acid sequence of most candidate loci matched the Wm82 (SCN-susceptible) sequence. Among candidate loci with amino acid substitutions, including *Glyma.07g19600* and *Glyma.07g196200* flanking *NSF/Glyma.07g195900* on the side not described in *SI Appendix,* Table 3, only NSF_{RAN07} encoded the same consistent amino acid changes across all examined *Rhg1*-containing germplasm.

A SNP associated with the Ch11 α-SNAP intron retention allele - a predicted SCN-resistance QTL – is also enriched among predicted *Rhg1*$^{+}$ accessions in the USDA collection

A recent study implicated the interval carrying the intron-retention allele of $\alpha\text{-SNAP}_{Ch11}$ ($\alpha\text{-}SNAP_{Ch11}\text{-}IR$) in SCN-resistance, but the responsible gene(s) within this QTL interval were not defined (Lakhssassi et al., 2017). The $\alpha\text{-}SNAP_{Ch11}\text{-}IR$ allele may have emerged randomly, or it may confer some selective advantage; for example, by reducing available levels of WT α-SNAP proteins and shifting the balance toward the disruptive *Rhg1* α-SNAP proteins. This could be

particularly relevant in *Rhg1*$_{LC}$ soybean lines that typically carry only three copies of *α-SNAP*$_{Rhg1LC}$ with correspondingly lower mRNA abundance, in contrast to the nine- or ten-copy *Rhg1*$_{HC}$ lines (Cook et al., 2012; Cook et al., 2014). We therefore used SoySNP50K data to analyze the frequency of the *α-SNAP*$_{Ch11}$*-IR* allele in the whole USDA collection and in the 855 *Rhg1*$^{+}$ *Glycine max* accessions noted above. We found a C/T SNP (ss715610416) located ~17,000 bp downstream of the *α-SNAP*$_{Ch11}$ locus that was associated with the *α-SNAP*$_{Ch11}$*-IR* allele, as indicated by our WGS data. Using immunoblots, we tested for total levels of WT α-SNAP protein among several *Rhg1*$_{LC}$ accessions that had either WT or *α-SNAP*$_{Ch11}$*-IR*-associated SNPs (ss715610416). The *Rhg1*$_{LC}$ accessions possessing the WT-linked SNP had higher WT α-SNAP abundance relative to the *Rhg1*$_{LC}$ accessions with the ss715610416 SNP (Fig. S10). Across the USDA soybean collection, we then found that the *α-SNAP*$_{Ch11}$*-IR*-associated ss715610416 genotype was present in 5.6% of accessions (Fig. 4*C*). Perhaps surprisingly, we observed the *α-SNAP*$_{Ch11}$ *IR*-associated ss715610416 genotype in only half (55.9%) of the *Rhg1*$_{LC}$ soybean lines and in about a third (34.7%) of the *Rhg1*$_{HC}$ lines (Fig. 4*D*). However, this enrichment of the *α-SNAP*$_{Ch11}$*-IR* linked-SNP within *Rhg1*$^{+}$ germplasm provides further evidence that this allele beneficially contributes to *Rhg1*$^{+}$ soybean varieties.

All *Rhg1*$^{+}$ recombinant inbred lines (RILs) derived from *Rhg1*$^{+}$ x *rhg1*$^{-}$ crosses inherit *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 QTL now known to encode *NSF*$_{RAN07}$ (Webb et al., 1995), we explored if

the *Rhg1*$^{+}$ progeny of more recent biparental crosses strictly inherited *NSF*$_{RAN07}$. From the Soybean Nested Association Mapping (SoyNAM) project (Song et al., 2017), we examined F5 genotypic data for populations of derived RILs developed from crosses of the IA3023 (SCN-susceptible) hub-parent to eight different soybean accessions carrying either *Rhg1*$_{HC}$ (seven accessions) or *Rhg1*$_{LC}$ (one accession). There were 122 to 139 RILs in each population. The segregation for *NSF*$_{RAN07}$: *NSF*$_{Ch07}$*WT* in soybean lines lacking *Rhg1* did not deviate from the null hypothesis of 1:1 segregation in six of the eight populations. 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 inherited *Rhg1*$_{HC}$ or *Rhg1*$_{LC}$ and a WT *NSF*$_{Ch07}$ allele. (Table S4). This was based upon the low-density SoySNP6K mapping data that used linked rather than perfect genetic markers for *Rhg1* and *NSF* (Song et al., 2017). We therefore genotyped all 8 of these RILs via sequencing and/or primers detecting the *Rhg1* repeat junction and a WT NSF$_{Ch07}$ vs. *NSF*$_{RAN07}$ allele and found that all 8 re-examined RILs containing *Rhg1*$_{HC}$ or *Rhg1*$_{LC}$ also carried the NSF$_{RAN07}$ ^$_{116}$F and M$_{181}$I mutations. Thus, all *Rhg1*$^{+}$ RILs also inherited *NSF*$_{RAN07}$ (Table S4). We analogously infer that the 5 lines of the Webb *et al.* study that appeared to break co-inheritance between *Rhg1*$_{HC}$ and *NSF*$_{RAN07}$ likely underwent a crossover between the RFLP markers linked to either *Rhg1* or *NSF* (Webb et al, 1995). Taken together with the biochemical and *in planta* impacts of *Rhg1* α-SNAPs and NSF$_{RAN07}$ described above, the SoySNP50K and NAM data indicate that *NSF*$_{RAN07}$ co-inheritance is a necessary balance that confers viability to soybeans carrying a resistance-type *Rhg1* haplotype.

2.5 Discussion

Across eukaryotes, NSF and α-SNAP interact through conserved electrostatic contacts to disassemble SNARE complexes, thereby maintaining cellular vesicle trafficking (Jahn and Scheller, 2006; Zhao and Brunger, 2016). Our findings indicate that *Rhg1*-mediated SCN resistance in soybean encompasses not just unusual changes in *Rhg1* α-SNAP sequence and abundance in syncytium cells, as previously published, but also changes in the other housekeeping α-SNAP and NSF genes whose products comprise the SNARE-recycling machinery. These results showcase how a functionally related set of multiple conserved housekeeping genes has co-evolved toward atypical forms, apparently to confer resistance to a highly damaging nematode pathogen while balancing plant fitness. The findings suggest that the two common resistance-conferring *Rhg1* haplotypes employ similar yet distinct strategies to combat SCN: they decrease WT α-SNAP availability via greater *Rhg1* copy number expansion and/or through loss of wild-type α-SNAP loci. We also found that presence of the unusual *Rhg1* α-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). Perhaps more importantly, this study and other recent work on *Rhg1* offer a molecular framework in which to understand the interactions of multiple QTLs associated with SCN resistance (Cook et al., 2014; Pant et al., 2014; Bekal et al., 2015; Bayless et al., 2016; Lakhssassi et al., 2017; Liu et al., 2017) : many of these loci modify the host vesicle fusion SNARE recycling machinery as a means of controlling SCN infection.

An understanding of the necessity of NSF_{RAN07} to balance *Rhg1* germplasm should become a central consideration in any planned transgenic addition of *Rhg1* into SCN-susceptible soybeans. Beyond soybean, this report suggests strategies to engineer *Rhg1*-like resistance into other cyst nematode-susceptible crop species, through introduction of sequence-edited α-SNAP alleles together with modulation of WT α-SNAP abundance and/or introduction of a compatible NSF.

It is biologically fascinating that complementary α-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. It highlights this pathway's importance during the pathogen-host interaction. The previous finding that *Rhg1* resistance-type α-SNAPs are impaired in normal NSF-interactions (Bayless et al., 2016) is supported by the present finding that a unique NSF allele - NSF_{RAN07} - is a requisite balance for *Rhg1* resistance-type α-SNAPs. While (Bayless et al., 2016) proposed the functional redundancy of multiple WT α-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_{RAN07} are apparently non-viable. Presence of WT α-SNAPs may still, in the presence of NSF_{RAN07}, contribute to the vigor and normal soybean yield of lines carrying the PI 88788 source of *Rhg1* ($Rhg1_{HC}$), but they are not sufficient to do so in the absence of NSF_{RAN07}.

Housekeeping genes have been reported to evolve particularly slowly due to selective constraints (Zhang and Li, 2004), which raises interest in the co-evolution between NSF_{RAN07} and *Rhg1* α-SNAP. It is unclear if existing natural variation at Ch07 NSF among soybean populations

enabled the development of the Ch18 *Rhg1* resistance-type α-SNAPs or vice versa, or if the *Rhg1* α-SNAP duplication event occurred first, followed by subsequent co-evolution of NSF and resistance-type α-SNAP polymorphisms. Currently, reports of natural NSF variation appear to be limited to humans. The 1,000 Human Genomes Project showed that in certain human ethnicities, NSF copy number expansions of up to 3 repeats are not uncommon (Sudmant et al., 2010). The original NSF locus is full length while the subsequent NSF copy number repeats truncate near exon 13 and do not encode full length NSF transcripts (Sudmant et al., 2010; Cabana-Dominguez et al., 2016). A recent study reported a correlation between this human NSF copy number variation and drug dependency (Cabana-Dominguez et al., 2016). Notably, no residue substitutions were reported among human NSF alleles, and to the best of our knowledge, no naturally occurring NSF protein variants from any organism have previously been reported.

As noted above, our findings about NSF_{RAN07} provide a mechanistic explanation for the previously observed segregation distortion, in SCN-resistant plants, between *Rhg1* and the chromosome 07 genetic interval that encodes NSF_{RAN07} (Webb et al., 1995; Kopisch-Obuch and Diers, 2006; Vuong et al., 2015). An observation that remains less firmly explained, however, is why transgenic expression of α-$SNAP_{Rhg1}$HC or α-$SNAP_{Rhg1}$LC protein, in *Agrobacterium rhizogenes*-transformed root systems of SCN-susceptible Williams 82 (which lacks NSF_{RAN07}), 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* α-SNAPs (Bayless et al., 2016). Notably, co-expression of $NSF_{N.benth}$ did not relieve the cell death in *N. benthamiana* leaves caused by *Rhg1* α-SNAPs, while WT soybean

NSF_{Ch07} did, albeit not as effectively as NSF_{RAN07}. Consistent with this, recombinant $NSF_{N.benth}$ essentially could not bind with *Rhg1* resistance-type α-SNAPs *in vitro* but those α-SNAPs could bind soybean WT NSF_{Ch07}. This may explain why soybean root cells do exhibit some tolerance of *Rhg1* α-SNAP expression even in the absence of NSF_{RAN07}. Nevertheless, the finding that all soybeans in the USDA collection that bear the signature of resistance-conferring *Rhg1* alleles also contain the NSF_{RAN07} R_4Q signature, coupled with the universal co-presence of the NSF_{RAN07} allele with *Rhg1* in the segregating progeny of NAM crosses, provides compelling evidence that at the organismal level, NSF_{RAN07} is essential for viability at some stage of growth for all *Rhg1*-containing germplasm.

$Rhg1_{LC}$ and *Rhg4* contribute together to the SCN resistance of $Rhg1_{LC}$ soybean lines (Liu et al., 2012; Mitchem, 2016) and it remains unclear why $Rhg1_{LC}$ confers only partial SCN resistance when *Rhg4*, which encodes a putative serine hydroxymethyltransferase, is absent (Brucker et al., 2005; Liu et al., 2012; Yu et al., 2016). Whether or not the *Rhg4* product directly impacts *Rhg1*-associated α-SNAP/NSF/SNARE interactions, consideration of the present findings may be influenced by published evidence that $Rhg1_{HC}$ soybean lines are substantially more effective than $Rhg1_{LC}^{+}$ $rhg4^{-}$ lines at conferring SCN resistance against HG type 0 SCN populations (Brucker et al., 2005; Yu et al., 2016)

The present findings add to what was already known or inferred about loss of some WT α-SNAPs in Peking-type $Rhg1_{LC}$ soybean lines (Matsye eta l., 2012; Cook et al., 2014; Lee et al., 2015; Lakhssassi et al., 2017). $Rhg1_{LC}$ varieties without or with the $\alpha\text{-}SNAP_{Ch11}\text{-}IR$ allele exhibit reduced or sharply reduced WT α-SNAP expression, respectively. This further supports the idea that, in addition to the unusual *Rhg1* α-SNAP proteins, WT α-SNAP levels and the [WT α-SNAP :

Rhg1 α-SNAP] ratio can be important determinants of successful *Rhg1*-mediated SCN resistance. Models for resistance involving evasion of nematode effectors should also be considered. NSF_{RAN07} may have allowed the non-toxic presence of *Rhg1*-type α-SNAPs, and *Rhg1* α-SNAPs may confer SCN resistance by failing to cooperate with nematode manipulation of the host. This model could explain why divergence of *Rhg1* α-SNAP types has occurred: different SCN populations may carry effectors that manipulate or interact with the host SNARE-recycling machinery, but to varying degrees depending on the α-SNAP protein that is present.

The *α-SNAP$_{Ch11}$ IR*-associated SNP, which correlated with modest changes in WT α-SNAP abundance, was present in only about half of the *Rhg1$_{LC}$* soybean lines and a third of the *Rhg1$_{HC}$* lines. Only a subtle positive impact on SCN resistance was reported for the broader QTL locus carrying the *α-SNAP$_{Ch11}$-IR* allele (Webb et al., 1995). However, because not all *Rhg1*$^{+}$ soybean lines carry the *α-SNAP$_{Ch11}$ IR*-associated genotype, its intentional use or exclusion may in the future translate to subtle but economically useful shifts in SCN resistance, in the HG type specificity of that resistance, or in soybean yield potential.

Discovery of the need for *NSF$_{RAN07}$* in *Rhg1*-containing soybeans may reveal a protective mechanism that reduces the toxic effects of *Rhg1* α-SNAPs in some cell types/conditions by facilitating participation of *Rhg1* α-SNAPs in productive 20S complexes that disassemble SNARE bundles, while the toxicity of *Rhg1* α-SNAPs remains predominant in syncytium cells. Such conditionally functional NSF mutants are known in the lab-derived *Drosophila* NSF *comatose* mutants, whereby the *comatose* allele encoded NSF-1 supports SNARE complex disassembly at room temperature but is non-functional at elevated temperatures, leading to failure of synaptic vesicle transport and fly paralysis (Littleton et al., 1998; Littleton et al., 2001). However, other

mechanistic hypotheses are viable. Future studies could examine the dynamics of NSF_{RAN07} abundance and function over time in developing SCN syncytia. For example, increased NSF levels were detected in syncytia in $Rhg1_{HC}$ varieties, and we had associated this with WT α-SNAP deficiency (Bayless et al., 2016), but whether it is NSF_{RAN07} or NSF_{Ch13} that increases is of interest and might suggest whether α-SNAP and NSF functionality is being promoted or disrupted by the host. We did observe that NSF_{RAN07} apparently can work with WT α-SNAPs, or at least is not toxic in the way that resistance-associated *Rhg1* α-SNAPs can be toxic. Expression of NSF_{RAN07} in *N. benthamiana* caused no macroscopically detectable leaf phenotypes, and NSF_{RAN07} is expressed in $Rhg1_{HC}$ soybeans that also express high levels of WT α-SNAPs. The random 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 NSF_{RAN07} likely functions effectively with WT α-SNAPs.

A summarizing model can be constructed. We hypothesize that co-expression of WT α-SNAPs or soybean NSFs can compete away the toxicity of *Rhg1* α-SNAPs by restoring functionally compatible partners to the 20S complex. The α-SNAPs bind bundles of three or four SNARE proteins and provide a key portion of the platform for binding NSF proteins and stimulation of ATP hydrolysis to disassemble those SNARE bundles. The success of the α-SNAP *N. benthamiana* toxicity assay apparently derives from the inability of $NSF_{N.benth}$ to function on SNARE bundles that carry *Rhg1* α-SNAPs. The phenotype caused by *Rhg1* α-SNAP expression is extreme in *N. benthamiana* but mild in most soybean cells because of the partial compatibility of *Rhg1* α-SNAPs with WT soybean NSFs. Our data indicate that even greater compatibility with *Rhg1* α-SNAPs is restored by presence of NSF_{RAN07}. Nevertheless, the findings of the present

work and ref. 17 indicate that *Rhg1* α-SNAPs are a less compatible partner than WT α-SNAPs. When the relative level of *Rhg1* α-SNAPs goes up, as has been documented for syncytium cells (Bayless et al., 2016), we hypothesize that the sub-optimal function of *Rhg1* α-SNAPs poisons syncytia. Alternative models for SCN resistance are possible; for example, the *Rhg1* α-SNAPs may be less sensitive to SCN effectors that manipulate WT α-SNAPs to the advantage of the nematode. In either case, we propose that NSF_{RAN07} is sufficiently compatible with *Rhg1* α-SNAPs to confer viability and productivity to *Rhg1*$^{+}$ soybean lines, especially when WT α-SNAPs are also abundant. NSF_{RAN07} may not be sufficient to overcome the toxicity of *Rhg1* α-SNAPs in syncytia. The lower abundance of WT α-SNAPs in many low-copy *Rhg1* lines may be important to enhancing the SCN resistance of those lines, where there are only three rather than ~10 tandem repeat copies of *Rhg1*, but it may also be a primary reason why low-copy *Rhg1* lines have been widely observed to exhibit minor reductions in grain yield.

The amassing evidence for the importance of altered α-SNAP/NSF/SNARE interactions in SCN-soybean interactions also suggests that these proteins may be attractive targets for cyst nematode effectors (Grunewald et al., 2009; Cook et al., 2014; Pant et al., 2014; Bekal et al., 2015; Bayless et al., 2016; Lakhssassi et al., 2017; Liu et al., 2017) . Preliminary evidence for one such effector is already in place (Bekal et al., 2015) and extensive variation is present in the SCN genes that encode putative SNARE-like protein effectors (Gardner et al., 2018). The gradual evolution of SCN populations toward an increasing number of individuals that can overcome the widely used *Rhg1*$_{HC}$ SCN resistance is a major issue for global soybean production (Niblack et al., 2008). 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 α-SNAP and NSF protein variants characterized in the present study.

2.6 Materials & Methods

Detailed information on procedures used for modeling, immunoblots, recombinant protein production, transgenic plant creation and care, plant crosses and SNP analysis are in the *SI Materials and Methods*.

In vitro α-SNAP NSF Binding Assays

In vitro α-SNAP NSF binding assays were performed with recombinant α-SNAP and NSF proteins essentially as described in (Barnard et al., 1996; Bayless et al., 2016). Briefly, 20 µg of recombinant α-SNAP was adhered to the bottom of a polypropylene tube at room temperature, and then washed. Subsequently, 20 µg of recombinant NSF was added to each tube containing immobilized α-SNAP and allowed to bind on ice for 10 mins, followed by 2 washes. α-SNAP and bound NSF were then collected by addition of 1X SDS-PAGE loading buffer and boiling, followed by separation on SDS-PAGE and visualization with silver staining.

2.7 Acknowledgements

We thank Dr. Shaojie Han for experiment contributions, and Drs. Sebastian Bednarek, Barry Ganetzky, Ann MacGuidwin and Matthew Hudson for additional guidance and suggestions. Thanks to Jaret Schroeder, Keilaa-Demi De La Cruz and Ryan Kessens for assistance in plant growth and care. This work was funded by United Soybean Board and Wisconsin Soybean Board

awards and by USDA-NIFA-AFRI award 2014-67013-21775 to A.F.B. The findings build on work supported by the National Science Foundation Graduate Research Fellowship under Grant No. (DGE-1256259) to A.M.B.

2.8 References

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.*

Bachem CW, et al. (2000) Antisense suppression of a potato alpha-SNAP homologue leads to alterations in cellular development and assimilate distribution. *Plant Mol Biol* 43(4):473-482.

Baker RW & Hughson FM (2016) Chaperoning SNARE assembly and disassembly. *Nat Rev Mol Cell Biol* 17(8):465-479.

Barnard RJ, Morgan A, & Burgoyne RD (1996) Domains of alpha-SNAP required for the stimulation of exocytosis and for N-ethylmalemide-sensitive fusion protein (NSF) binding and activation. *Mol Biol Cell* 7(5):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(4):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(5):599-603.

Bayless AM, et al. (2016) Disease resistance through impairment of alpha-SNAP-NSF interaction and vesicular trafficking by soybean Rhg1. *Proc Natl Acad Sci U S A* 113(47):E7375-E7382.

Bekal S, et al. (2015) A SNARE-Like Protein and Biotin Are Implicated in Soybean Cyst Nematode Virulence. *PLoS One* 10(12):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(1):44-49.

Cabana-Dominguez J, et al. (2016) A Highly Polymorphic Copy Number Variant in the NSF Gene is Associated with Cocaine Dependence. *Sci Rep* 6:31033.

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(3):264-270.

Concibido VC, Diers BW, & Arelli PR (2004) A Decade of QTL Mapping for Cyst Nematode Resistance in Soybean. *Crop Sci*. 44(4):1121-1131.

Cook DE, et al. (2012) Copy number variation of multiple genes at Rhg1 mediates nematode resistance in soybean. *Science* 338(6111):1206-1209.

Cook DE, et al. (2014) Distinct Copy Number, Coding Sequence, and Locus Methylation Patterns Underlie Rhg1-Mediated Soybean Resistance to Soybean Cyst Nematode. *Plant Physiol* 165(2):630-647.

Dodds PN & Rathjen JP (2010) Plant immunity: towards an integrated view of plant-pathogen interactions. *Nat Rev Genet* 11(8):539-548.

Gardner M, et al. (2018) Novel global effector mining from the transcriptome of early life stages of the soybean cyst nematode Heterodera glycines. *Sci Rep* 8(1):2505.

Gheysen G & Mitchum MG (2011) How nematodes manipulate plant development pathways for infection. *Curr Opin Plant Biol* 14(4):415-421.

Goodstein DM, et al. (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res* 40(Database issue):D1178-1186.

Grant D, Nelson RT, Cannon SB, & Shoemaker RC (2010) SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Res* 38(Database issue):D843-846.

Griff IC, Schekman R, Rothman JE, & Kaiser CA (1992) The yeast SEC17 gene product is functionally equivalent to mammalian alpha-SNAP protein. *The Journal of biological chemistry* 267(17):12106-12115.

Grunewald W, Cannoot B, Friml J, & Gheysen G (2009) Parasitic nematodes modulate PIN-mediated auxin transport to facilitate infection. *PLoS Pathog* 5(1):e1000266.

Hanson PI & Whiteheart SW (2005) AAA+ proteins: have engine, will work. *Nat Rev Mol Cell Biol* 6(7):519-529.

Hewezi T & Baum TJ (2013) Manipulation of plant cells by cyst and root-knot nematode effectors. *Mol Plant Microbe Interact* 26(1):9-16.

Horsnell WG, Steel GJ, & Morgan A (2002) Analysis of NSF mutants reveals residues involved in SNAP binding and ATPase stimulation. *Biochemistry* 41(16):5230-5235.

Jahn R & Scheller RH (2006) SNAREs--engines for membrane fusion. Nat Rev Mol Cell Biol 7(9):631-643.

Jones JT, et al. (2013) Top 10 plant-parasitic nematodes in molecular plant pathology. *Mol Plant Pathol* 14(9):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(2):199-207.

Kyndt T, Vieira P, Gheysen G, & de Almeida-Engler J (2013) Nematode feeding sites: unique organs in plant roots. *Planta* 238(5):807-818.

Lakhssassi N, et al. (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(8):1774-1791.

Littleton JT, et al. (1998) Temperature-sensitive paralytic mutations demonstrate that synaptic exocytosis requires SNARE complex assembly and disassembly. *Neuron* 21(2):401-413.

Littleton JT, et al. (2001) SNARE-complex disassembly by NSF follows synaptic-vesicle fusion. *Proc Natl Acad Sci U S A* 98(21):12233-12238.

Liu S, et al. (2012) A soybean cyst nematode resistance gene points to a new mechanism of plant resistance to pathogens. *Nature* 492(7428):256-260.

Liu S, et al. (2017) The soybean GmSNAP18 gene underlies two types of resistance to soybean cyst nematode. *Nat Commun* 8:14822.

Matsye PD, et al. (2012) The expression of a naturally occurring, truncated allele of an alpha-SNAP gene suppresses plant parasitic nematode infection. *Plant Mol Biol* 80(2):131-155.

Miao Y, et al. (2013) An essential and NSF independent role for alpha-SNAP in store-operated calcium entry. *Elife* 2:e00802.

Mitchum MG, et al. (2013) Nematode effector proteins: an emerging paradigm of parasitism. *New Phytol*.

Mitchum MG (2016) Soybean Resistance to the Soybean Cyst Nematode Heterodera glycines: An Update. *Phytopathology* 106(12):1444-1450.

Naydenov NG, et al. (2012) Loss of soluble N-ethylmaleimide-sensitive factor attachment protein alpha (alpha SNAP) induces epithelial cell apoptosis via down-regulation of Bcl-2 expression and disruption of the Golgi. *The Journal of biological chemistry* 287(8):5928-5941.

Niblack TL, et al. (2002) A Revised Classification Scheme for Genetically Diverse Populations of Heterodera glycines. *J Nematol* 34(4):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.

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.

Pant SR, et al. (2014) Syntaxin 31 functions in Glycine max resistance to the plant parasitic nematode Heterodera glycines. *Plant Mol Biol* 85(1-2):107-121.

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(3):1241-1253.

Sanderfoot AA, Assaad FF, & Raikhel NV (2000) The Arabidopsis genome. An abundance of soluble N-ethylmaleimide-sensitive factor adaptor protein receptors. *Plant Physiol* 124(4):1558-1569.

Sanyal S & Krishnan KS (2001) Lethal comatose mutation in Drosophila reveals possible role for NSF in neurogenesis. *Neuroreport* 12(7):1363-1366.

Severin AJ, et al. (2010) RNA-Seq Atlas of Glycine max: A guide to the soybean transcriptome. *BMC Plant Biology* 10(1):160.

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 Q, et al. (2013) Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. *PLoS One* 8(1):e54985.

Song Q, et al. (2015) Fingerprinting Soybean Germplasm and Its Utility in Genomic Research. *G3* (Bethesda) 5(10):1999-2006.

Song Q, et al. (2017) Genetic Characterization of the Soybean Nested Association Mapping Population. *Plant Genome* 10(2).

Sudmant PH, et al. (2010) Diversity of human copy number variation and multicopy genes. *Science* 330(6004):641-646.

Vuong TD, et al. (2015) Genetic architecture of cyst nematode resistance revealed by genome-wide association study in soybean. *BMC Genomics* 16:593.

Webb DM, et al. (1995) Genetic mapping of soybean cyst nematode race-3 resistance loci in the soybean PI 437.654. *Theor Appl Genet* 91(4):574-581.

Whiteheart SW, Schraw T, & Matveeva EA (2001) N-ethylmaleimide sensitive factor (NSF) structure and function. *Int Rev Cytol* 207:71-112.

Whiteheart SW & Matveeva EA (2004) Multiple binding proteins suggest diverse functions for the N-ethylmaleimide sensitive factor. *J Struct Biol* 146(1-2):32-43.

Wickner W & Schekman R (2008) Membrane fusion. *Nat Struct Mol Biol* 15(7):658-664.

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.

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(12):2403-2412.

Zhang L & Li WH (2004) Mammalian housekeeping genes evolve more slowly than tissue-specific genes. *Mol Biol Evol* 21(2):236-239.

Zhao C, Slevin JT, & Whiteheart SW (2007) Cellular functions of NSF: not just SNAPs and SNAREs. *FEBS Lett* 581(11):2140-2149.

Zhao C, Matveeva EA, Ren Q, & Whiteheart SW (2010) Dissecting the N-ethylmaleimide-sensitive factor: required elements of the N and D1 domains. *The Journal of biological chemistry* 285(1):761-772.

Zhao M, et al. (2015) Mechanistic insights into the recycling machine of the SNARE complex. *Nature* 518(7537):61-67.

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(9 Pt B):1912-1926.

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(18):E2290-2297.

2.9 Figures

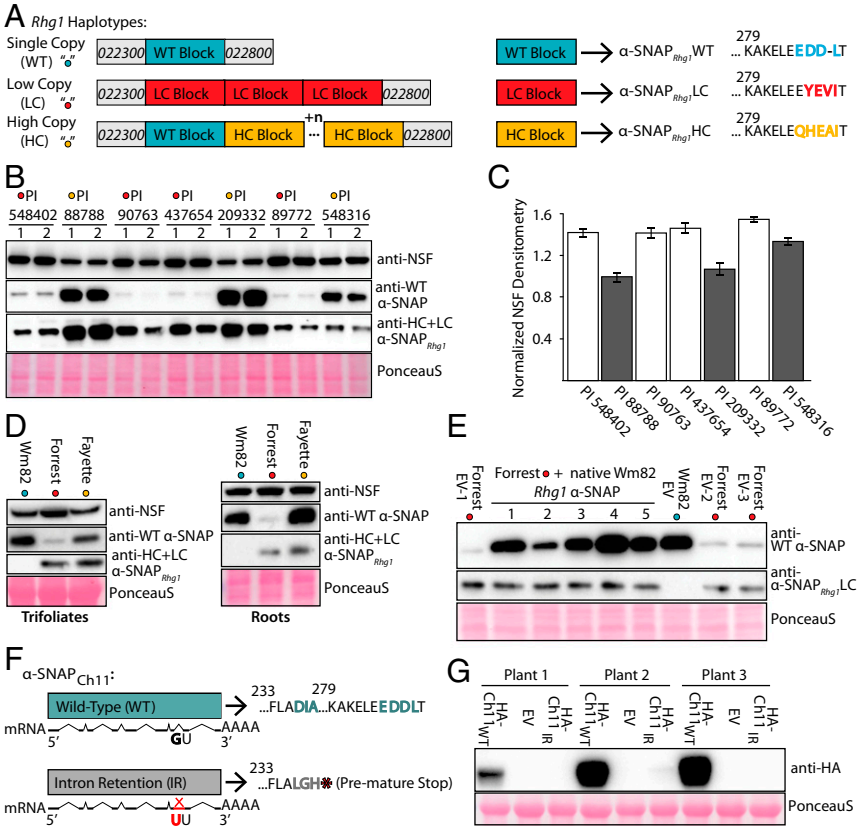


Fig. 1. WT α-SNAP proteins are much less abundant while NSF is more abundant in *Rhg1*$_{\text{LowCopy}}$ soybeans. (A) Schematic of *Rhg1* haplotype classes. (Left) *Rhg1* WT (shown blue), *Rhg1* LC (shown red), *Rhg1* HC (shown orange; n = variable HC-type repeat numbers); not drawn to scale. The C-terminal amino acid poly- morphisms encoded by the *Rhg1* α-SNAPs are shown at Right. HC *Rhg1* haplotypes retain a single WT-like *Rhg1* repeat. (B) Immunoblot of WT α-SNAPs, *Rhg1* resistance-type α-SNAPs and NSF in roots of soybean HG test varieties (two samples for each genotype). *Rhg1*$_{\text{LC}}$ varieties (red dot; 3 Rhg1 copies): PI 548402 (Peking), PI 89772, PI 437654, PI 90763; *Rhg1*$_{\text{HC}}$ varieties (orange dot): PI 88788 (9 copies), PI 209332 (10 copies), PI

548316 (7 copies). PonceauS staining shows similar loading of total protein. (C) Densitometry indicating total NSF expression in HG type test lines. (D) Like B, but immunoblots for trifoliate leaves or roots of Wm82 and modern *Rhg1*$_{LC}$ and *Rhg1*$_{HC}$ varieties Forrest and Fayette. (E) Immunoblots for total WT α-SNAPs and α-SNAP$_{Rhg1}$LC in Forrest (*Rhg1*$_{LC}$) transgenic roots transformed with an empty vector (EV; three transgenic lines) or with the native Wm82 α-SNAP$_{Rhg1}$WT locus (five transgenic lines), or in WT Wm82 roots transformed with EV. (F) Schematic of chromosome 11 α-SNAP alleles with exon/intron models, and nucleotide and amino acid polymorphisms. (G) The encoded α-SNAP$_{Ch11}$ intron retention protein, unlike the WT α-SNAP$_{Ch11}$, does not accumulate. Anti-HA immunoblot of total protein from *N. benthamiana* leaves is agroinfiltrated to express empty vector, N-HA-α-SNAP$_{Ch11}$, or N-HA-α-SNAP$_{Ch11}$-IR (intron retention). PonceauS staining shows similar loading of total protein.

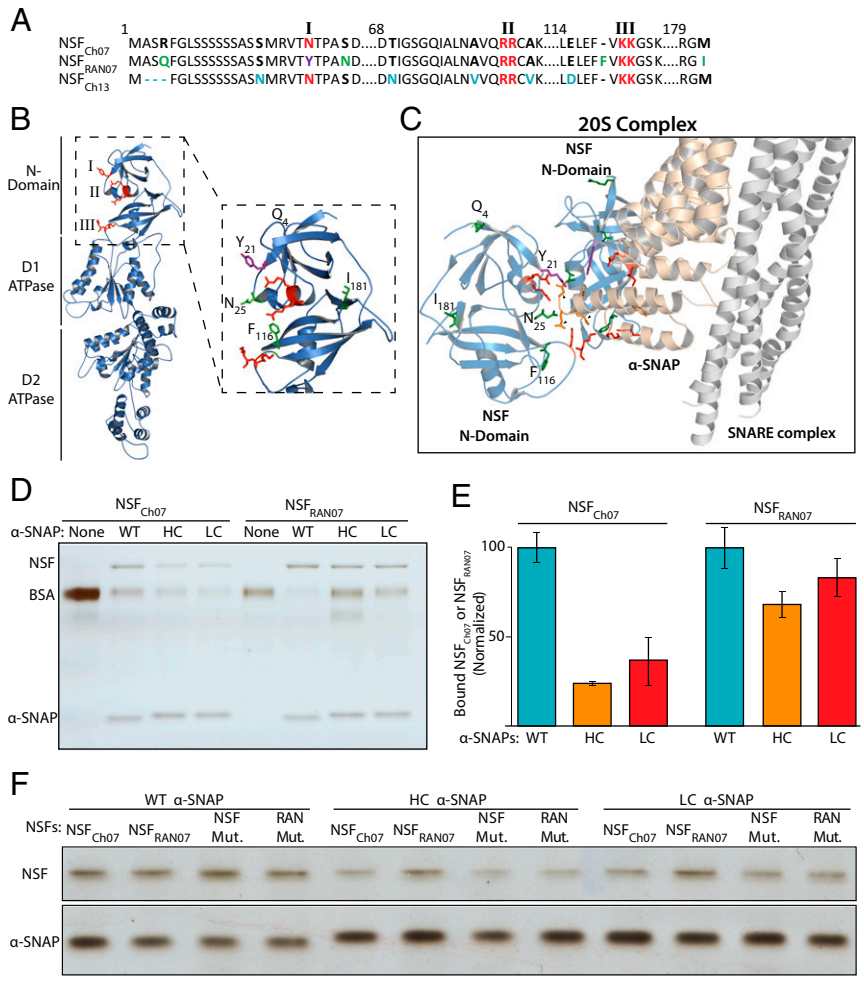


Fig. 2. Rhg1-containing lines carry an *NSF$_{Ch07}$* allele (*RAN07*) with N-domain polymorphisms at the α-SNAP binding interface that enhance binding with polymorphic *Rhg1* resistance-type α-SNAPs. (A) Alignment of N-terminal domains of soybean NSF$_{Ch07}$, NSF$_{Ch13}$, and NSF$_{RAN07}$. Large identical regions are omitted. N-domain residues corresponding to those that bind α-SNAP are colored red (N_{21}, $RR_{82–83}$, $KK_{117–118}$). NSF$_{RAN07}$ polymorphisms R_4Q, $S_{25}N$, $_{116}F$, and $M_{181}I$ are colored green or purple ($N_{21}Y$); unique NSF$_{Ch13}$ residues are colored light blue. (B) NSF$_{RAN07}$ modeled to NSF$_{CHO}$ cryo-EM structure (3J97A, State II). NSF residue patches implicated in α-SNAP binding are colored red and labeled I, II, or III. Zoomed-in view shows NSF$_{RAN07}$ N-domain polymorphisms colored green or purple ($N_{21}Y$). (C) Cryo- EM structure of mammalian 20S

supercomplex, masked to show only SNARE bundle (white), one α-SNAP (yellow), and two NSF N domains (light blue). Shown are the mammalian residues; conserved NSF N-domain patches (I, R_{10}; II, $RK_{67\text{-}68}$; III, $KK_{104–105}$) are shown in red, and α-SNAP C-terminal contacts ($D_{217}DEED_{290–293}$) are shown in orange. Black arrowheads point to three orange α-SNAP residues $EED_{291–293}$ corresponding to sites of C-terminal polymorphisms in α-$SNAP_{Rhg1}$HC and α-$SNAP_{Rhg1}$LC. NSF_{RAN07} polymorphism sites are colored green, except $N_{21}Y$ is in purple. (D) Silver-stained SDS/PAGE showing amount of recombinant NSF_{Ch07} or NSF_{RAN07} bound in vitro by a fixed quantity of the recombinant α-SNAP protein indicated on second line: no-α-SNAP control (None) or WT, LC or HC *Rhg1* α-SNAP. (E) Densitometric quantification of NSF_{Ch07} or NSF_{RAN07} bound as in D by the *Rhg1* α-SNAPs denoted at bottom; data are from three independent experiments, and error bars show SEM. (F) Like D, but showing recombinant NSF_{Ch07}, NSF_{RAN07}, or mutants of either, bound in vitro by *Rhg1* α-SNAPs. NSF Mut. and RAN Mut. refer to NSF_{Ch07} $N_{21}A$ $F_{115}A$ and NSF_{RAN07} $Y_{21}N$ F_{116}^, respectively.

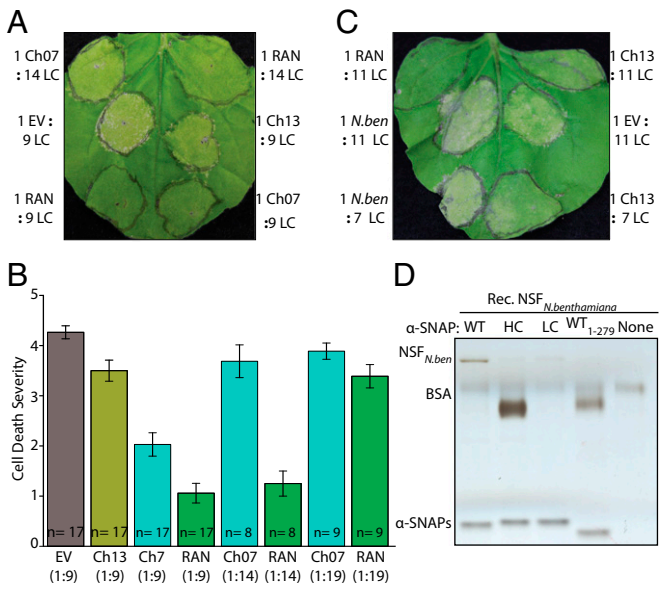


Fig. 3. Coexpression of soybean NSFs reduces cell death symptoms caused by α-SNAP$_{Rhg1}$LC; NSF$_{RAN07}$ gives strongest protection. (A) *N. benthamiana* leaves ~6 d after agroinfiltration with 9:1 or 14:1 strain mixture (9 or 14 parts *Agrobacterium* that delivers LC [α-SNAP$_{Rhg1}$LC] to one part *Agrobacterium* that delivers the indicated soybean NSF [NSF$_{Ch07}$, NSF$_{RAN07}$, NSF$_{Ch13}$] or EV control). (B) Scoring of cell death severity, across multiple independent experiments, in *N. benthamiana* leaf patches coexpressing NSF$_{Ch07}$, NSF$_{RAN07}$, or NSF$_{Ch13}$; n is number of leaves scored; error bars show SEM. (C) Like A, but 7:1 or 11:1 mixed cultures expressing α-SNAP$_{Rhg1}$LC with NSF$_{RAN07}$, NSF$_{N.benth}$, NSF$_{Ch13}$, or empty vector. (D) Silver-stained SDS/PAGE of recombinant NSF$_{N.benth}$ bound in vitro by recombinant WT, LC, or HC *Rhg1* α-SNAP proteins or WT α-SNAP lacking the final 10 C-terminal residues (WT$_{1–279}$).

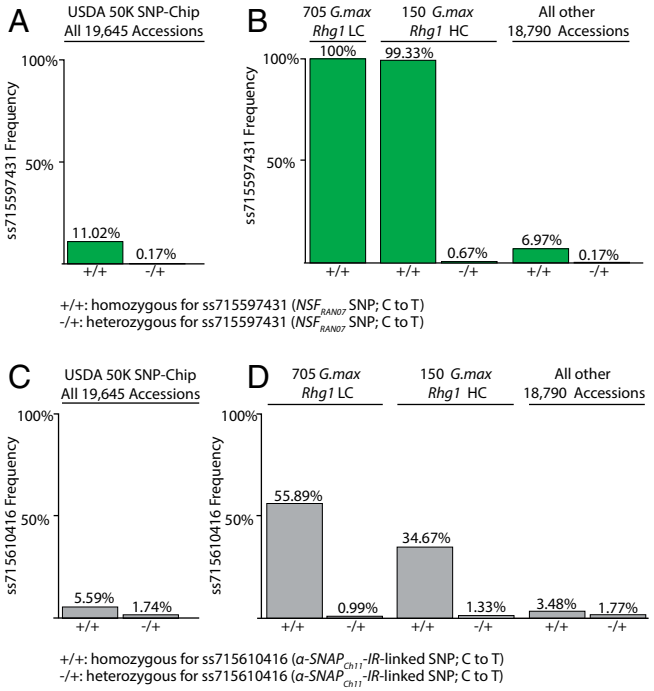


Fig. 4. All soybeans in the USDA germplasm collection that carry an *Rhg1+* SNP signature also carry the R_4Q NSF_{RAN07} polymorphism. (A) Frequency of SoySNP50K SNP ss715597431 (corresponding to NSF_{RAN07} R_4Q) in all 19,645 SoySNP50K-genotyped *Glycine max* accessions. (B) Frequency of ss715597431 in all USDA collection *G. max* with $Rhg1_{LC}$ or $Rhg1_{HC}$ haplotype signatures, or in the remainder of SoySNP50K-genotyped *G. max*. (C) Frequency of SoySNP50K SNP ss715610416 that is closest marker for α-$SNAP_{Ch11}$-IR allele, across all 19,645 genotyped USDA *G. max* accessions. (D) Frequency of ss715610416 in all USDA collection *G. max* with $Rhg1_{LC}$ or $Rhg1_{HC}$ haplotype signatures, or in remainder of SoySNP50K-genotyped *G. max*.

2.10 Supporting Information

Recombinant Protein Production

Vectors encoding recombinant α-SNAP$_{Rhg1}$HC, α-SNAP$_{Rhg1}$LC, α-SNAP$_{Rhg1}$WT, α-SNAP$_{Rhg1}$WT$_{1-279}$ and the WT alleles of NSF *Glyma.07G195900* (NSF$_{Ch07}$) and *Glyma.13G180100* (NSF$_{Ch13}$) were generated in Bayless *et al*., 2016. The open reading frames (ORFs) encoding the soybean NSFRAN07 allele of *Glyma.07G195900* or *N. benthamiana* NSF were cloned into the expression vector pRham N-His-SUMO according to manufacturer instructions (Lucigen). Recombinant α-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_{600} ~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-α-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 α-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 flow-through. Recombinant protein purity was assessed by Coomassie blue staining and quantified via a spectrophotometer.

***In vitro* NSF-α-SNAP Binding Assays**

In vitro NSF binding assays were performed essentially as described in (Barnard et al., 1996; Bayless et al., 2016). Briefly, 20 µg of each respective recombinant α-SNAP protein was added to the bottom of a 1.5-mL polypropylene tube and incubated at 25°C for 20 min. Unbound α-SNAP proteins were then washed by adding α-SNAP wash buffer [25 mM Tris, pH 7.4, 50 mM KCl, 1 mM DTT, 0.1 mg/mL bovine serum albumin (BSA)]. After removal of wash buffer, 20 µg of recombinant NSF (1 µg/µL 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 α-SNAP was then calculated using densitometric analysis with ImageJ.

Antibody Production and Validation

Affinity-purified polyclonal rabbit antibodies raised against α-SNAP$_{Rhg1}$HC, α-SNAP$_{Rhg1}$LC and wild-type α-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 α-SNAP residues: “EEDDLT,” “EQHEAIT,” or “EEYEVIT” for wild-type, high-, or low-copy α-SNAPs, respectively. For NSF, a synthetic peptide, “ETEKNVRDLFADAEQDQRTRGDESD,” corresponding to residues 300 to 324 of the *Glyma.07G195900* encoded protein was used. This same epitope is also present in *Glyma.13G180100* encoded NSF$_{Ch13}$ and this NSF antibody was also 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_2(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* α-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 from cotyledons of the noted genetic background 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_{600} 0.60 for NSF constructs or OD_{600} 0.80 for α-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 $MgCl_2$, 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−2·s−1 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_{RAN07} or the *N. benthamiana* NSF were PCR amplified from a root cDNA library of *Rhg1*$_{LC}$ variety, "Forrest" or a *N. benthamiana* leaf cDNA library using KAPA HiFi polymerase, respectively. Expression cassettes for $NSF_{N.benthamiana}$, $NSF_{Ch13,}$ NSF_{Ch07} and NSF_{RAN07} ORFs were directly assembled into a pBluescript vector containing the previously described soybean ubiquitin (GmUbi) promoter and NOS terminator using Gibson assembly (Cook et al., 2012). 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-$SNAP_{Ch11}$ Intron-Retention (IR) allele was amplified with Kapa HiFi from a root cDNA library of *Rhg1*$_{LC}$ variety "Forrest" while the ORF encoding WT α-$SNAP_{Ch11}$ was previously generated in (Cook et al., 2012). The ORFs encoding either α-$SNAP_{Ch11}$ and α-$SNAP_{Ch11}$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-

$SNAP_{Rhg1}$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-$SNAP_{Ch11}$ was amplified from gDNA of Williams82 in two separate fragments (3.25 kb and 3.60 kb fragments) and Gibson assembled into the binary pSM101 vector cut with BamHI-PstI.

Segregating NAM Crosses

Soybean parental crosses and 6K SNP genotyping mapping were developed and performed by (Song et al., 2017).

Protein Structure Modeling

NSF_{RAN07}, α-$SNAP_{Ch11}$ and α-$SNAP_{Ch11}$IR structural homology models were generated using SWISS-MODEL and the resulting PDB files were analyzed with PyMol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC). NSF_{RAN07} was modeled to NSF_{CHO} (*Cricetulus griseus*, Chinese hamster ovary) (PDB 3j97.1) cryo-EM structure from Zhao *et al* (Brunger group). 20S supercomplex modeling was also generated using PDB 3j97, with α-SNAPs and SNARE complexes (VAMP-2, Syntaxin-1A, SNAP-25) of *Rattus norvegicus* origin (Zhao et al., 2015). α-$SNAP_{Ch11}$ and α-$SNAP_{Ch11}$IR were modeled to sec17 (yeast α-SNAP) crystal structure 1QQE donated courtesy of Rice *et al* (Brunger group)(Rice and Brunger, 1999).

Sequence Logo and Alignments

The R_4Q NSF amino acid consensus logo was generated using the first 10 NSF amino acids of the model eukaryotic organisms using WebLogo (Crooks et al., 2004). The NSF amino acid sequences of these organisms were retrieved from publicly available sequence data at the National Center for Biotechnology Information (NCBI). Plant NSF sequences were obtained from Phytozome.org and aligned using Jalview (Waterhouse et al., 2009).

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 Williams82 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).

Primers used in the study

N. ben NSF Rev TCAATATCGAGCAATGTCCTGA

N. ben NSF For	ATGGCAGGGAGATTTGGTTCC
N. ben Gmubi For	GATTTATCTGTGATTGTTGACTCGACAGATGGCAGGGAGATTTGGTTCC
N. ben NSF NOS Rev	GAAAGCTGGGTCGAATTCGCCCTTTTCAATATCGAGCAATGTCCTGAAGGC
N. ben pRham For	CACCGCGAACAGATTGGAGGTGCAGGGAGATTTGGTTCC
N. ben pRham Rev	GTGGCGGCCGCTCTATTATCAATATCGAGCAATGTCCTGA
NSF Ch07 UTR For	GCCATTGCTATTGTGGTGCGA
NSF Ch07 UTR Rev	CTATCAGCACAACCAAACAAACACTG
RAN07 Gmubi For	GATTTATCTGTGATTGTTGACTCGACAGATGGCGAGTCAGTTCGGG
RAN07 NOS Rev	AGCTGGGTCGAATTCGCCCTTTTCATAACCTAACAACATCCTGGAGGC
RAN07 pRham For	CTCACCGCGAACAGATTGGAGGTGCGAGTCAGTTCGGGTTATCG
RAN07 pRham Rev	CAGCGGTGGCGGCCGCTCTATTATAACCTAACAACATCCTGGAGGC
RAN07 Detect For	CTACACGCCCGCGAACGAC
NSF07 Detect Rev	CTCACTTGTACGGAATCACCGG
WT NSF Detect For	CAACACGCCCGCGAGCGAC
NSF 13 UTR For	GCCAAGAAACAGAGAAACATAGAGGC
NSF 13 UTR Rev	CTGAACAGTAACAAGCAATGTAGGAATG
NSF 13 NOS Rev	GAAAGCTGGGTCGAATTCGCCCTTTTcATCTAACAACATCCTGGAGGCAATC
NSF 13 Gmubi For	GATTTATCTGTGATTGTTGACTCGACAGATGTTCGGCTTATCGTCTTCGTCTTCC -TC
Ch11 Native 1 For	CTAATCTGGGGACCTGGGTACCCGGGCTCGAACACGTATAAAGGACCTGAGG
Ch11 Native 1 R	GGATCAATATCTCGAAGCTGAATATGGGC
Ch11 Native 2 For	CCCATATTCAGCTTCGAGATATTGATCC

Ch11 Native 2 Rev	GGGCCCACAGATATTAATTAAGACATCTGCAGCATCTCTCTTTATTTCCATG
	-CACGCG
Ch11 UTR For	CGATCAATCCATCCATCTTCACTTGC
Ch11 UTR Rev	CAAACAATAGGTCCAACCGCCAG
Ch18 Native For	CCTGCAGGGAGCAGTAGGCTTCTTTGGAACTTG
Ch18 Native Rev	GTTTGCCACTTTAGGAACCCTAGG

SI References

Barnard RJ, Morgan A, & Burgoyne RD (1996) Domains of alpha-SNAP required for the stimulation of exocytosis and for N-ethylmalemide-sensitive fusion protein (NSF) binding and activation. *Mol Biol Cell* 7(5):693-701.

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(47):E7375-E7382.

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(6111):1206-1209.

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(2):630-647.

Crooks GE, Hon G, Chandonia JM, & Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14(6):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(5):491-498.

Goodstein DM, et al. (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res* 40(Database issue):D1178-1186.

Li H & Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754-1760.

Rice LM & Brunger AT (1999) Crystal structure of the vesicular transport protein Sec17: implications for SNAP function in SNARE complex disassembly. *Mol Cell* 4(1):85-95.

Severin AJ, et al. (2010) RNA-Seq Atlas of Glycine max: A guide to the soybean transcriptome. *BMC Plant Biology* 10(1):160.

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 genetics* 11(5):e1005200.

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(2).

Waterhouse AM, Procter JB, Martin DM, Clamp M, & Barton GJ (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25(9):1189-1191.

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(7537):61-67.

SI Figures

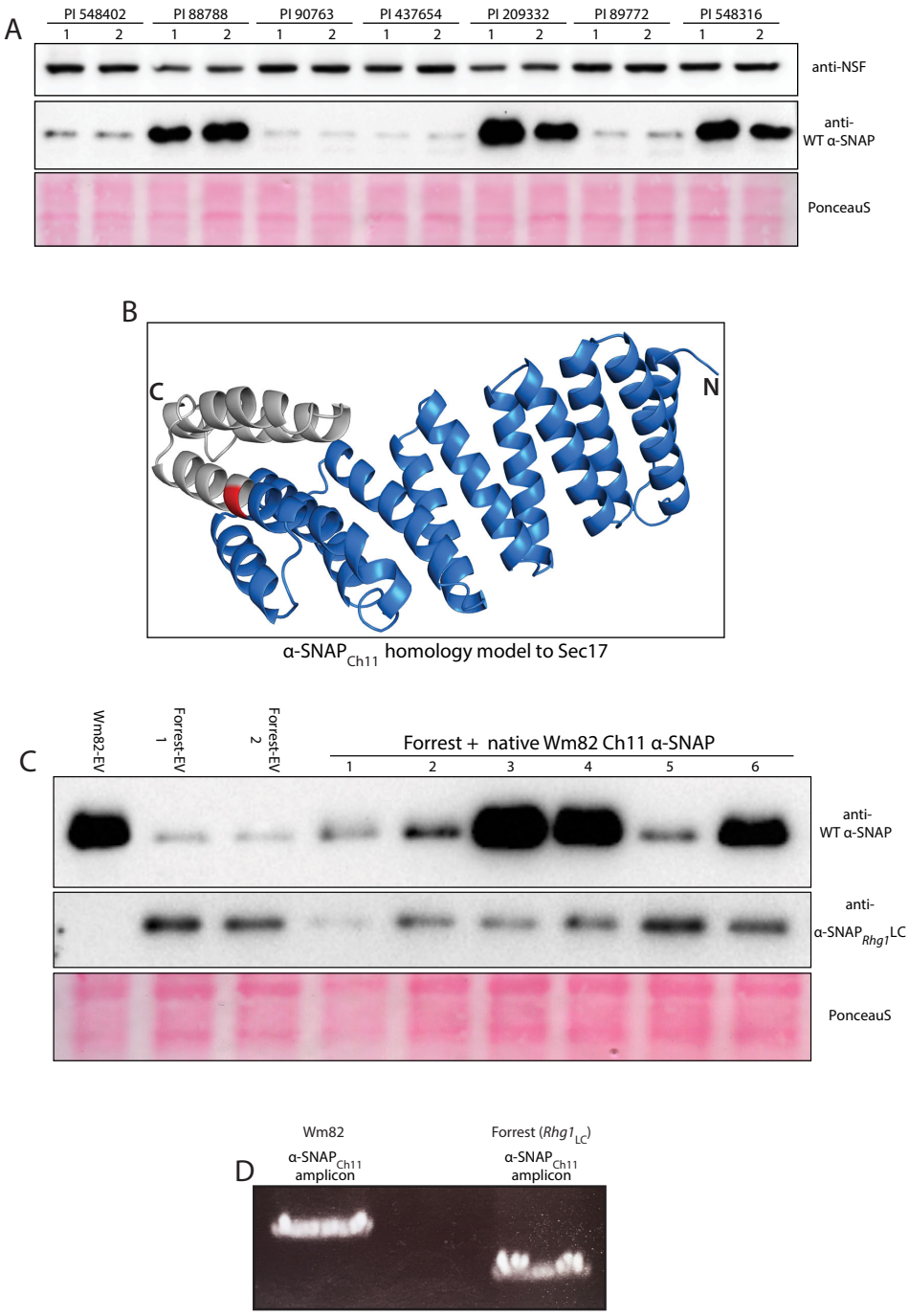


Fig. S1. Wild-type α-SNAP expression is reduced in *Rhg1*$_{\text{Low Copy}}$ soybeans. (*A*) Independent immunoblot like Fig. 1B and incorporated into NSF densiometric analyses shown in Fig. 1*C*.

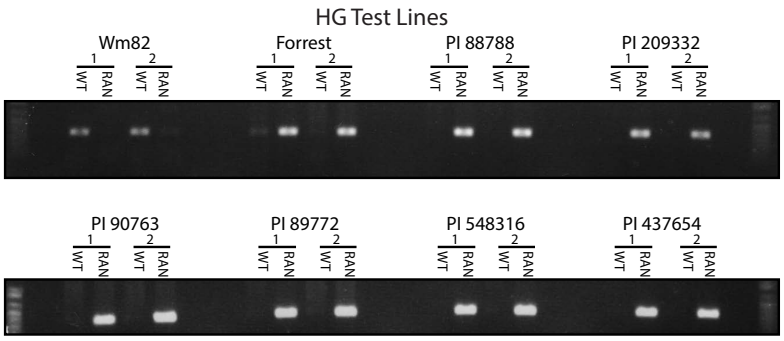
Immunoblot of wild-type α-SNAPs and NSF expression in HG-Type soybean roots. *Rhg1*$_{LC}$ varieties: PI 548402 (Peking), PI 89772, PI 437654, PI 90763; *Rhg1*$_{HC}$ varieties: PI 88788, PI 209332, PI 548316 (7 copy). PonceauS staining shows total protein loaded per lane. (*B*) Modeling of α-SNAP$_{Ch11}$ to Sec17 crystal structure (yeast α-SNAP, PDB ID 1QQE) suggests early termination of alpha-helix 12 in the intron-retention mutant. Pre-mature translational termination point shown red. (*C*) Immuno- blots for total WT α-SNAP and α-SNAP$_{Rhg1}$LC levels in Forrest (*Rhg1*$_{LC}$) transgenic roots transformed with the native WT α-SNAP$_{Ch11}$ locus from Wm82 or an EV (empty vector) control. (*D*) Agarose gel showing PCR amplicons of the promoter regions of the α-SNAP$_{Ch11}$-IR allele or the WT α-SNAP$_{Ch11}$ allele from Wm82.

NSF RAN07 alignment to Wild-Type NSF_{Ch07} (Wm82)

A

```
Wms82           MASRFGLSSSSSSASSMRVTNTPASDLALTNLAFCSPSDLRNFAVPGHNNLYLAAVADSF
RAN07           MASQFGLSSSSSSASSMRVTYTPANDLALTNLAFCSPSDLRNFAVPGHNNLYLAAVADSF
                ***:*************** ***.**********************************

Wms82           VLSLSAHDTIGSGQIALNAVQRRCAKVSSGDSVQVSRFVPPEDFNLALLTLELEF VKKGS
RAN07           VLSLSAHDTIGSGQIALNAVQRRCAKVSSGDSVQVSRFVPPEDFNLALLTLELEFFVKKGS
                ************************************************** *****

Wms82           KSEQIDAVLLAKQLRKRFMNQVMTVGQKVLFEYHGNNYSFTVSNAAVEGQEKSNSLERGM
RAN07           KSEQIDAVLLAKQLRKRFMNQVMTVGQKVLFEYHGNNYSFTVSNAAVEGQEKSNSLERGI
                ************************************************************:

Wms82           ISDDTYIVFETSRDSGIKIVNQREGATSNIFKQKEFNLQSLGIGGLSAEFADIFRRAFAS
RAN07           ISDDTYIVFETSRDSGIKIVNQREGATSNIFKQKEFNLQSLGIGGLSAEFADIFRRAFAS
                ************************************************************

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

Wms82           TEKNVRDLFADAEQDQRTRGDESDLHVIIFDEIDAICKSRGSTRDGTGVHDSIVNQLLTK
RAN07           TEKNVRDLFADAEQDQRTRGDESDLHVIIFDEIDAICKSRGSTRDGTGVHDSIVNQLLTK
                ************************************************************

Wms82           IDGVESLNNVLLIGMTNRKDMLDEALLRPGRLEVQVEISLPDENGRLQILQIHTNKMKEN
RAN07           IDGVESLNNVLLIGMTNRKDMLDEALLRPGRLEVQVEISLPDENGRLQILQIHTNKMKEN
                ************************************************************

Wms82           SFLAADVNLQELAARTKNYSGAELEGVVKSAVSYALNRQLSLEDLTKPVEEENIKVTMDD
RAN07           SFLAADVNLQELAARTKNYSGAELEGVVKSAVSYALNRQLSLEDLTKPVEEENIKVTMDD
                ************************************************************

Wms82           FLNALHEVTSAFGASTDDLERCRLHGMVECGDRHKHIYQRAMLLVEQVKVSKGSPLVTCL
RAN07           FLNALHEVTSAFGASTDDLERCRLHGMVECGDRHKHIYQRAMLLVEQVKVSKGSPLVTCL
                ************************************************************

Wms82           LEGSRGSGKTALSATVGIDSDFPYVKIVSAESMIGLHESTKCAQIIKVFEDAYKSPLSVI
RAN07           LEGSRGSGKTALSATVGIDSDFPYVKIVSAESMIGLHESTKCAQIIKVFEDAYKSPLSVI
                ************************************************************

Wms82           ILDDIERLLEYVPIGPRFSNLISQTLLVLLKRLPPKGKKLMVIGTTSELDFLESIGFCDT
RAN07           ILDDIERLLEYVPIGPRFSNLISQTLLVLLKRLPPKGKKLMVIGTTSELDFLESIGFCDT
                ************************************************************

Wms82           FSVTYHIPTLNTTDAKKVLEQLNVFTDEDIDSAAEALNDMPIRKLYMLIEMAAQGEHGGS
RAN07           FSVTYHIPTLNTTDAKKVLEQLNVFTDEDIDSAAEALNDMPIRKLYMLIEMAAQGEHGGS
                ************************************************************

Wms82           AEAIFSGKEKISIAHFYDCLQDVVRL
RAN07           AEAIFSGKEKISIAHFYDCLQDVVRL
                *************************
```

B



C

Gene	Young leaf	Flower	One cm pod	Pod shell 10DAF	Pod shell 14DAF	Seed 10DAF	Seed 14DAF	Seed 21DAF	Seed 25DAF	Seed 28DAF	Seed 35DAF	Seed 42DAF	Root	Nodule
Glyma.07G195900	5	8	8	9	9	3	4	6	5	2	4	2	11	7
Glyma.13G180100	6	6	8	6	5	3	4	3	2	1	3	1	12	7

Normalized RNA-seq reads from Wm82 (soybase.org)

Fig. S2. The *NSF_{RAN07}* allele is present within all examined *Rhg1* HG-Type test lines. (*A*) NSF_{RAN07} amino acid alignment with NSF_{Ch07} of soybean reference genome Williams 82 (Wm82). N-domain amino acid polymorphisms unique to NSF_{RAN07} shown red. Corresponding residues of Wm82 encoded NSF_{Ch07} (wild-type) shown boldface. (*B*) Agarose gel showing PCR amplicons generated with *NSF_{RAN07}* (RAN) or *NSF_{Ch07}* WT (WT) allele specific primers on the HG-Type soybeans or soybean Wm82. (*C*) Wm82 normalized RNA-seq reads for both NSF_{Ch07} and NSF_{Ch13} across soybean tissues. RNA-seq data from Severin *et al* (2010) and this RNA-seq atlas data is publicly available at Soybase.org. DAF: days after fertilization.

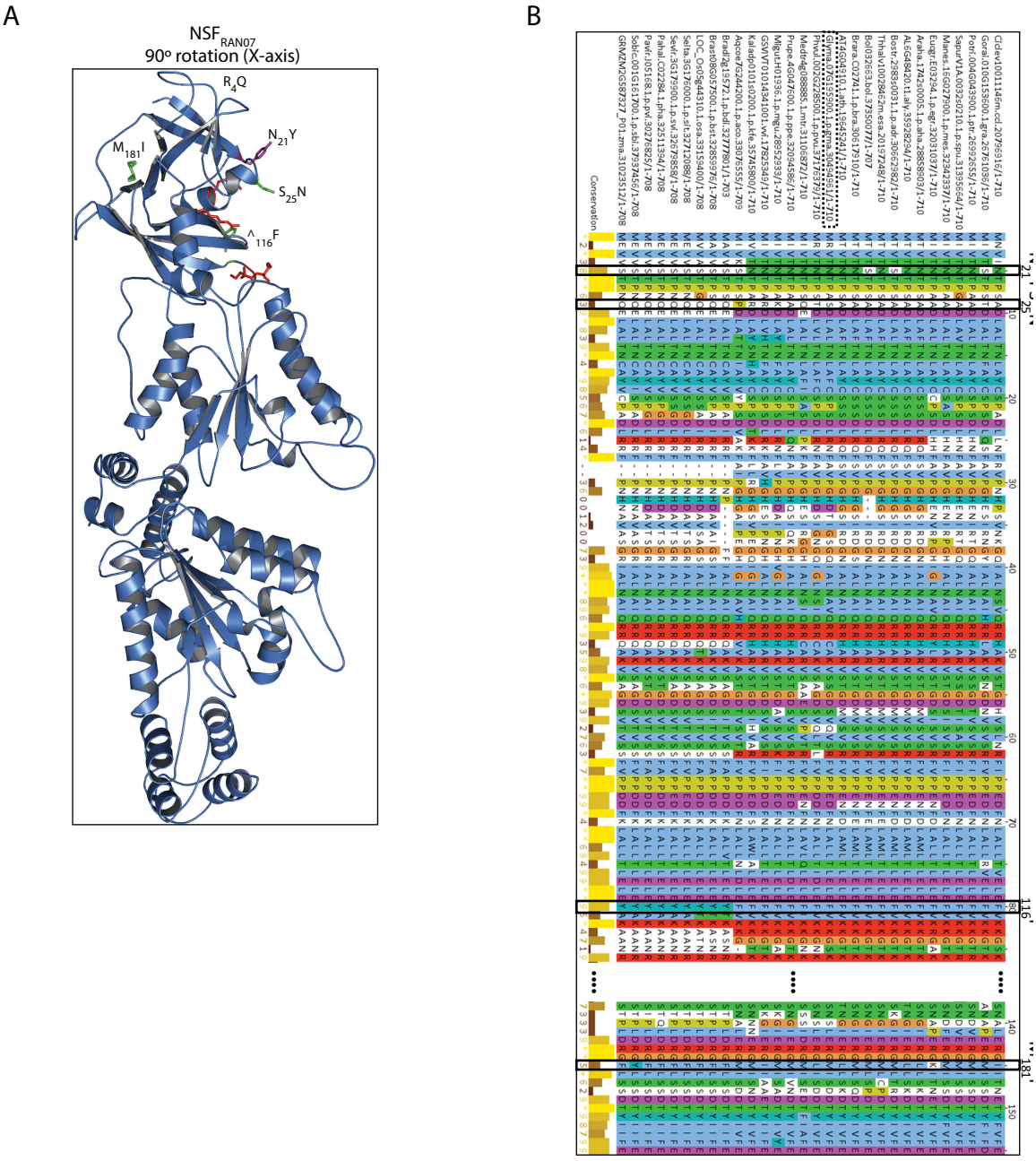
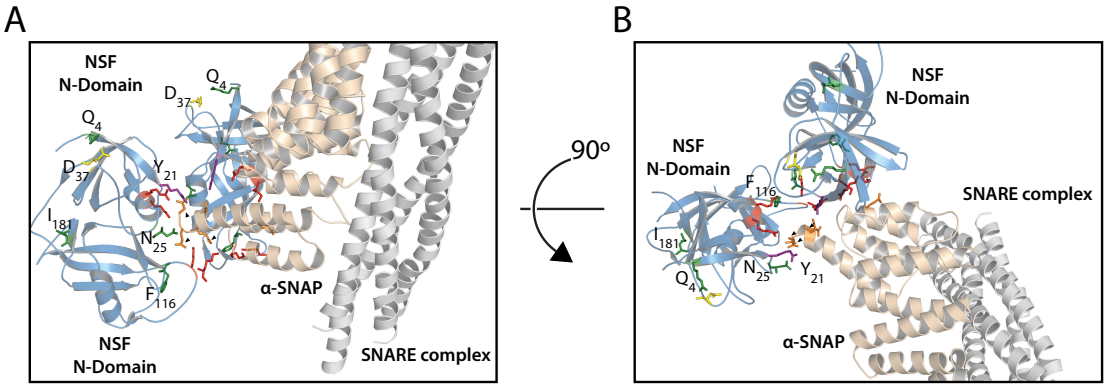
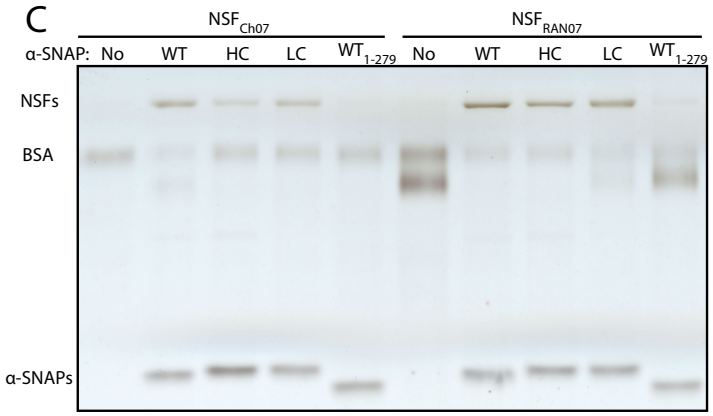
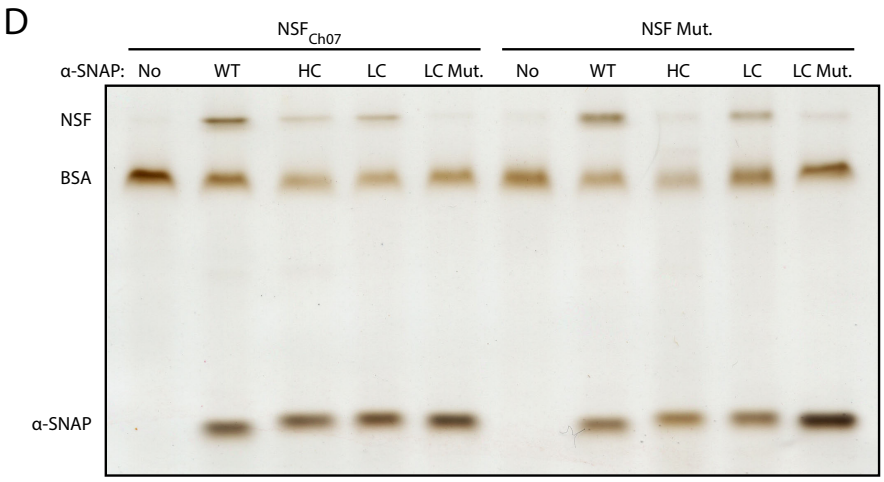


Fig. S3. (*A*) NSF_{RAN07} modeled to NSF_{CHO} cryo-EM structure as in Fig. 2A, but rotated 90° on X-axis relative to Fig. 2*B*. NSF residue patches implicated in α-SNAP binding colored red and labeled I, II or III, respectively. (*B*) Alignment of NSF N-domain using available plant NSF amino acid sequences from Phytozome.org (Goodstein et al., 2012). Alignment generated with Jalview starting at a conserved methionine residue corresponding to NSF_{RAN07} methione 17. Residues

polymorphic in NSF_{RAN07} are outlined with a box with the corresponding NSF_{RAN07} polymorphism/position labeled above. "...." indicates a gap of residues not polymorphic in NSF_{RAN07}.


A
NSF
N-Domain
Q4
D37
Q4
D37
Y21
I181
N25
F116
α-SNAP
NSF
N-Domain
SNARE complex
90°
B
NSF
N-Domain
NSF
N-Domain
F116
SNARE complex
I181
Q4
N25
Y21
α-SNAP


C
NSFCh07
NSFRAN07
α-SNAP: No WT HC LC WT1-279 No WT HC LC WT1-279
NSFs
BSA
α-SNAPs


D
NSFCh07
NSF Mut.
α-SNAP: No WT HC LC LC Mut. No WT HC LC LC Mut.
NSF
BSA
α-SNAP

E

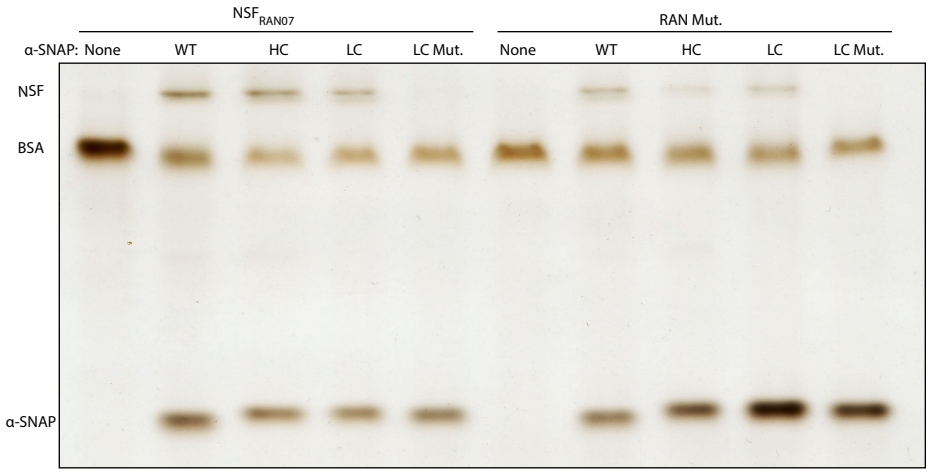


Fig. S4. NSF$_{RAN07}$ polymorphisms are at the α-SNAP binding interface. NSF$_{RAN07}$ and NSF$_{Ch07}$ binding with α-SNAP is dependent on α-SNAP C-terminal polymorphisms, and two NSF$_{RAN07}$ polymorphisms enhance binding by *Rhg1* resistance type α-SNAPs. (*A*) Like Fig. 2*C*, cryo-EM structure of mammalian 20S supercomplex showing SNARE bundle (white), one α-SNAP (yellow) and two NSF N-domains (light blue). Conserved NSF N-domain patches (R_{10}; RK_{67-68}; $KK_{104-105}$) shown red, α-SNAP C-terminal contacts ($D_{217}DEED_{290-293}$) shown orange and α-SNAP residues corresponding to *Rhg1* polymorphisms indicated by black arrows, NSF$_{RAN07}$ polymorphisms (R_4Q, $S_{25}N$, $\wedge_{116}F$, $M_{181}I$, ^=insertion) colored green, except polymorphisms $N_{21}Y$ colored in purple. NSF$_{RAN07}$ polymorphism R_4Q positions near an acidic residue D_{37} (shown yellow). (*B*) Same as A, but rotated 90° on Y-axis. (*C*) Same as Fig. 2*D*, except recombinant NSF$_{Ch07}$ or NSF$_{RAN07}$ bound *in vitro* by no-α-SNAP (*D,E*). Silver-stained SDS-PAGE showing amount of NSF$_{Ch07}$, NSF$_{RAN07}$, or mutants of either (”NSF Mut.”, “RAN Mut.”) bound to constant amount of *Rhg1* α-SNAPs, including α-SNAP$_{Rhg1}$LC $286_{AAAA}289$ (”LC Mut.”), which has alanine substitutions at the *Rhg1* polymorphisms . NSF Mut. is NSF$_{Ch07}N_{21}A$ $F_{115}A$; RAN Mut. is NSF$_{RAN07}$

$Y_{21}N\ F_{116}$^. WT, HC or LC refers to α-SNAP$_{Rhg1}$WT, α-SNAP$_{Rhg1}$HC or α-SNAP$_{Rhg1}$LC, while “None” is a no α-SNAP negative binding control. Entirely independent replicate binding experiments were performed as in C, D and E with similar results.

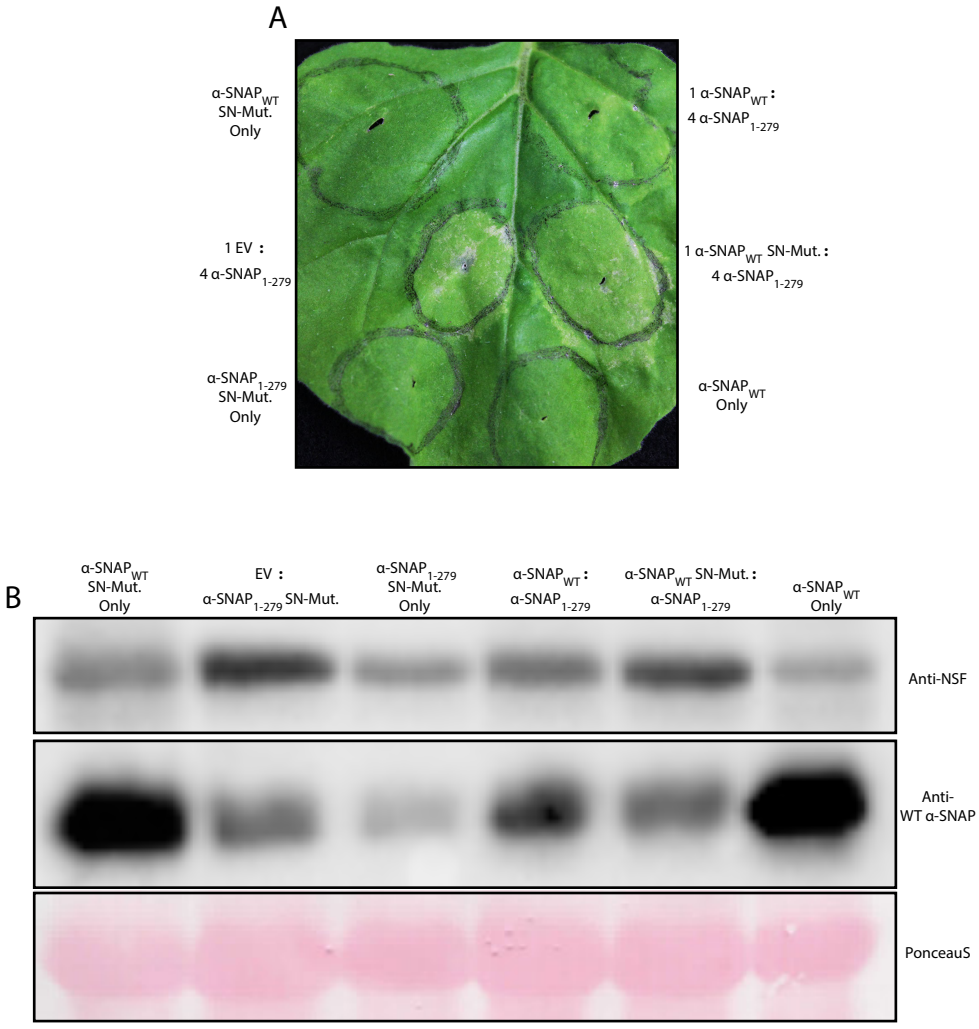


Fig. S5. Soybean WT α-SNAPs mutated at residues known in other α-SNAPs to promote SNARE-bundle interactions are deficient in rescuing the *N. benthamiana* cell death induced by toxic α-SNAP types, and α-SNAP$_{1-279}$ (which lacks the final 10 C-terminal residues and induces *N. benthamiana* cell death) becomes unable to cause cell death when mutated at the same SNARE-bundle interaction residues . (*A*) Representative *N. benthamiana* leaf infiltrated with individual or mixed Agrobacterium cultures expressing the indicated α-SNAP constructs. The α-SNAP SNARE-bundle interaction mutant (”SN-Mut.”) is $K_{193}E$ $R_{230}E$, as described in Zhao *et al*

(2015). α-SNAP$_{WT}$ is the WT (SCN-susceptible Williams 82) chromosome 18 α-SNAP$_{Rhg1}$WT. (*B*) Immunoblot of total WT α-SNAP or NSF proteins in *N. benthamiana* leaves expressing only the indicated solo proteins, or construct mixes. PonceauS staining indicates similar levels of total protein loading.

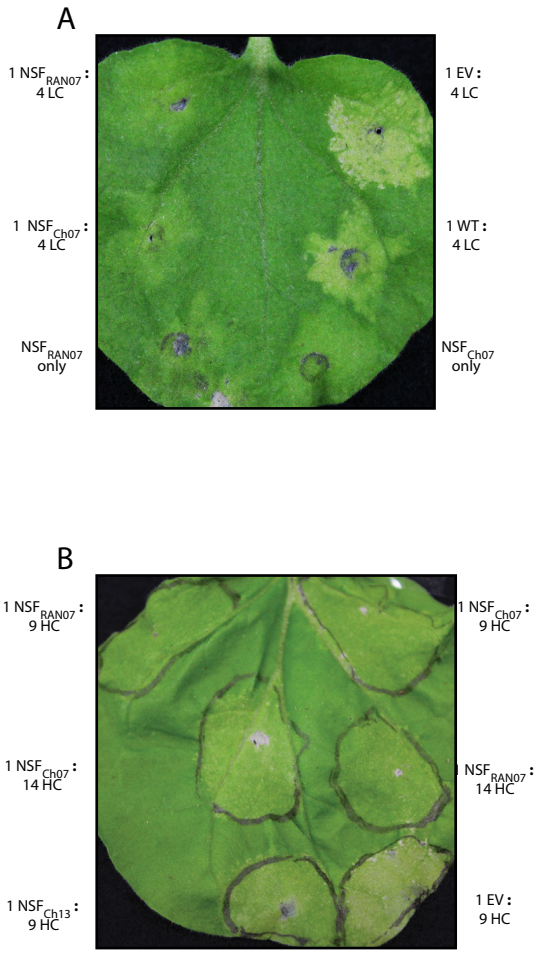


Fig. S6. Coexpression of soybean NSFs reduces cell-death symptoms caused by *Rhg1* resistance α-SNAPs; NSF_{RAN07} gives strongest protection. (*A*) *N. benthamiana* leaves ~6 days post agro-infiltration with 1:4 mixed cultures of NSF_{Ch07} or NSF_{RAN07} or α-$SNAP_{Rhg1}$WT or empty vector to α-$SNAP_{Rhg1}$LC (four parts *Agrobacterium* delivering α-$SNAP_{Rhg1}$LC to one part *Agrobacterium* delivering a soybean NSF, or α-$SNAP_{Rhg1}$WT or empty vector control). (*B*) Like Fig. 3*A*, but using α-$SNAP_{Rhg1}$HC instead of α-$SNAP_{Rhg1}$LC in the corresponding mixture cultures of NSF_{Ch07} or NSF_{RAN07} or empty vector.

```
Ch07            MASRFGLSSSSSSASSMRVTNTPASDLALTNLAFCSPSDLRNFAVPGHNNLYLAAVADSF
Nben            MAGRFG-----SGASTMIVTNTPAKDLAYTNCAYCSPADLRNFLVPGSK-LAYGLIADAF
                          *.**:* ******.*** ** *:***:***** *** : *  . :**:*

Ch07            VLSLSAHDTIGSGQIALNAVQRRCAKVSSGDSVQVSRFVPPEDFNLALLTLELEFVKKGS
Nben            VLTLAAHDGIPNGHLGLNAIQRRYAKVSTGDTISVNRFVPPDDFNLALLTIDLEFVKKGT
                **:*:*** * .*::.***:*** ****:**::.*.*****:********::*******:

Ch07            KSEQIDAVLLAKQLRKRFMNQVMTVGQKVLFEYHGNNYSFTVSNAAVEGQEKSNSLERGM
Nben            RDEQVDAVSLANQVRKKFANQIMSTGQKVTFEYHGNSYIFTVNQATVEGQEKSN-IERGM
                :.**:*** **:*:**:* **:*:.**** ******.* ***.:*:******** :****

Ch07            ISDDTYIVFETSRDSGIKIVNQREGATSNIFKQKEFNLQSLGIGGLSAEFADIFRRAFAS
Nben            ISADTYIIFEAANSSGIKIVNQREAASSSIFRQKEFNLQSLGIGGLSAEFADIFRRAFAS
                ** ****:**::..**********.*:*.**:*****************************

Ch07            RVFPPHVTSKLGIKHVKGMLLYGPPGTGKTLMARQIGKILNGKEPKIVNGPEVLSKFVGE
Nben            RVFPPHVTSKLGIKHVKGMLLYGPPGTGKTLMARQIGKMLNGKEPKIVNGPEVLSKFVGE
                ************************************:*********************

Ch07            TEKNVRDLFADAEQDQRTRGDESDLHVIIFDEIDAICKSRGSTRDGTGVHDSIVNQLLTK
Nben            TEKNVRDLFADAEQDQRTKGDQSELHVIIFDEIDAICKSRGSTRDGTGVHDSIVNQLLTK
                ******************:**:*:************************************

Ch07            IDGVESLNNVLLIGMTNRKDMLDEALLRPGRLEVQVEISLPDENGRLQILQIHTNKMKEN
Nben            IDGVESLNNVLLIGMTNRKDLLDEALMRPGRLEVQVEISLPDENGRLQILQIHTNQMKEN
                *********************:*****:****************************:****

Ch07            SFLAADVNLQELAARTKNYSGAELEGVVKSAVSYALNRQLSLEDLTKPVEEENIKVTMDD
Nben            SFLSPDVNLQELAARTKNYSGAELEGVVKSAVSFALNRQLSMDDLTKPVDEESIKVTMDD
                ***:.*****************************:*******::******:**.*******

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

Ch07            LEGSRGSGKTALSATVGIDSDFPYVKIVSAESMIGLHESTKCAQIIKVFEDAYKSPLSVI
Nben            LEGPSGSGKTAMAATVGIESDFPYVKIISAETMIGLSESSKCAQIVKVFEDAYKSPLSIV
                ***. ******::*****:********:***:**** **:*****:************::

Ch07            ILDDIERLLEYVPIGPRFSNLISQTLLVLLKRLPPKGKKLMVIGTTSELDFLESIGFCDT
Nben            VLDGIERLLEYVAIGPRFSNLISQTLLVLLKRLPPKGKKILVIGTTSEAGFLDSVGLCDA
                :**.********.****************************::******* .**:*:*:**:

Ch07            FSVTYHIPTLNTTDAKKVLEQLNVFTDEDIDSAAEALNDMPIRKLYMLIEMAAQGEHGGS
Nben            FSVTYHVPTLKTEDAKKVLQQLNVFSNDDVDSAAEALNDMPIKKLYMVVEMAAQGEHGGT
                ******:***:* ******:*****:::*:************:****::**********:

Ch07            AEAIFSGKEKISIAHFYDCLQDVVRL
Nben            AEAIYSGKEKIQISHFYDCLQDIARY
                ****:******.*:********:.*
```

Fig. S7. Amino acid sequence of NSF clone from *N. benthamiana* aligned with NSF_{Ch07} from soybean Williams82. NSF N-domain residues conserved in α-SNAP binding are shown red in boldface.

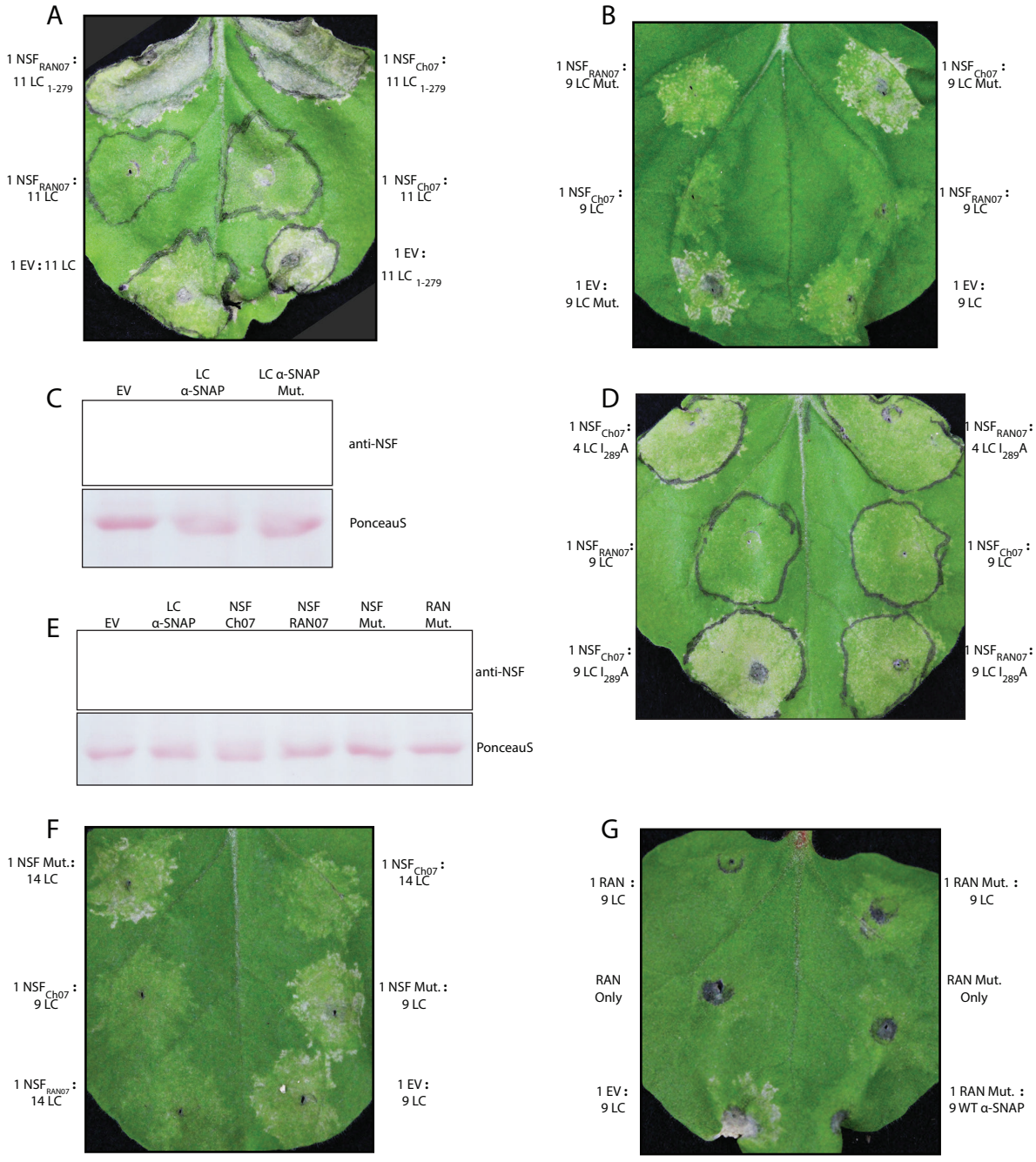


Fig. S8. Coexpression of soybean NSFs reduces cell-death symptoms caused by *Rhg1* resistance-type α-SNAPs; changes to the polymorphic α-SNAP$_{Rhg1}$LC C-terminus reduce cell death protection, as do site-directed mutations at implicated NSF residues. (*A*) Like Fig. 3A, *N. benthamiana* mixed culture of α-SNAP$_{Rhg1}$LC or α-SNAP$_{Rhg1}$LC$_{1-279}$ Ch07 or NSF$_{RAN07}$ or empty

vector. (*B*) A 9:1 mixed culture of α-SNAP$_{Rhg1}$LC or α-SNAP$_{Rhg1}$LC 286$_{AAAA}$289 (”LC Mut.”) co-expressed with NSF$_{Ch07}$ or NSF$_{RAN07}$ or empty vector. (*C*) Immunoblot of total NSF protein expression from *N. benthamiana* leaves expressing empty vector (EV), α-SNAP$_{Rhg1}$LC or α-SNAP$_{Rhg1}$LC 286$_{AAAA}$289 (LC Mut.). PonceauS staining indicates similar loading of total proteins. (*D*) Like A and B, but 4:1 or 9:1 mixed cultures of α-SNAP$_{Rhg1}$LC or α-SNAP$_{Rhg1}$LC-I$_{289}$A co-expressed with NSF$_{Ch07}$ or NSF$_{RAN07}$. (*E*) Immunoblot of total NSF protein expression from *N. benthamiana* leaf tissues expressing empty vector (EV), α-SNAP$_{Rhg1}$LC, or the indicated NSF constructs. PonceauS staining indicates similar loading of total proteins. (*F*) A 9:1 mixed culture of α-SNAP$_{Rhg1}$LC co-expressed with either EV, NSF$_{Ch07}$, or NSF$_{RAN07}$ or NSF$_{Ch07}$ N$_{21}$A F$_{115}$A (NSF Mut.). (*G*) A 9:1 mixed culture of α-SNAP$_{Rhg1}$LC co-expressed with either EV, NSF$_{RAN07}$ or NSF$_{RAN07}$ Y$_{21}$N F$_{116^{\wedge}}$ (RAN Mut.).


10
20
30
40
50
Cre10.g461250.t1.1.cre.30790752/1-719
Vocar.0029s0014.1.p.vca.32887558/1-719
192553.mpu.27345663/1-736
30476.olu.27413522/1-737
mrna30162.1-v1.0-hybrid.fve.27255095/1-725
AL7G31640.t1.aly.35937077/1-741
AL7G50070.t1.aly.35937513/1-733
Araha.11265s0006.1.p.aha.28838070/1-726
AHYPO_011645-RA.ahy.32837333/1-721
GSMUA_Achr9P05850_001.mac.32305992/1-743
GSMUA_Achr8P27650_001.mac.32309832/1-747
Pavir.Eb03112.1.p.pvi.30310572/1-736
Sevir.5G318800.1.p.svi.32673667/1-724
Seita.5G315900.1.p.sit.32683068/1-724
Brast08G057500.1.p.bst.32859976/1-745
GRMZM2G587327_P01.zma.31023512/1-741
Sevir.3G179900.1.p.svi.32679858/1-743
Seita.3G176000.1.p.sit.32712088/1-743
Pavir.J05168.1.p.pvi.30276825/1-741
Pahal.C02284.1.pha.32511394/1-741
evm_27.model.AmTr_v1.0_scaffold00068.142.atr.31573525/1-725
Aqcoe7G244200.1.p.aco.33076555/1-747
Aqcoe4G103100.1.p.aco.33066341/1-745
Medtr4g088885.1.mtr.31106872/1-752
Glyma.13G180100.1.p.gma.30504626/1-742
Phvul.002G228500.1.p.pvu.37176379/1-744
Migut.H01936.1.p.mgu.28952933/1-736
AT4G04910.1.ath.19645241/1-742
AL6G48420.t1.aly.35928294/1-740
Araha.1742s0005.1.p.aha.28858903/1-740
Bostr.2983s0031.1.p.adr.30662982/1-740
Carubv10002710m.cru.20907336/1-721
Cagra.5070s0007.1.p.cgr.28901595/1-769
Thhalv10028462m.esa.20197248/1-740
Bol032663.bol.37350077/1-736
Brara.I02193.1.p.bra.30641152/1-740
Bol030577.bol.37348896/1-714
Brara.C02741.1.p.bra.30617910/1-740
Kalax.0433s0008.1.p.kla.32568442/1-744
Kaladp0024s0326.1.p.kfe.35770053/1-746
Kalax.0066s0054.1.p.kla.32566782/1-731
Kaladp0101s0200.1.p.kfe.35745800/1-744
Kalax.0610s0020.1.p.kla.32561630/1-744
Kalax.0588s0024.1.p.kla.32597742/1-744
Lus10011177.lus.23147134/1-753
Lus10015440.lus.23160937/1-733
mrna02390.1-v1.0-hybrid.fve.27273079/1-729
GSVIVT01014341001.vvi.17825349/1-739
Eucgr.E03298.1.p.egr.32029837/1-763
Ciclev10011146m.ccl.20796916/1-744
orange1.1g004550m.csi.18122541/1-745
Cucsa.180440.1.csa.16967312/1-743
Prupe.4G047600.1.p.ppe.32094586/1-744
MDP0000228175.m.22632077/1-737
MDP0000853659.m.22672616/1-712
evm.model.supercontig_27.27.cpa.16415645/1-753
Eucgr.E03294.1.p.egr.32031037/1-743
29851.m002421.rco.16811448/1-701
Gorai.010G153600.1.gra.26761036/1-743
Thecc1EG029866t2.tca.27439713/1-719
SapurV1A.0012s0300.1.p.spu.31433784/1-740
Potri.004G043900.1.ptr.26992655/1-740
SapurV1A.0032s0210.1.p.spu.31395664/1-740
Manes.16G027900.1.p.mes.32342337/1-740
Manes.17G045800.1.p.mes.32365918/1-740
Pp3c2_19730V3.1.p.ppa.32934002/1-742
Sphfalx0054s0105.1.p.sfa.32620768/1-737
e_gw1.11.372.1.smo.15413278/1-725
e_gw1.158.23.1.smo.15417735/1-729
Mapoly0104s0033.1.p.mpo.33012542/1-770
Conservation

Fig. S9. Alignment of available plant NSF sequences starting at predicted residue 1. General consensus of R_4 is observed across a majority of plant species. Alignment generated with Jalview using all available angiosperm NSF sequences from Phytozome.org (Goodstein et al., 2012). Only NSF sequences of residue lengths comparable to known NSF sequences (~700-800 residues) were used for the alignment.

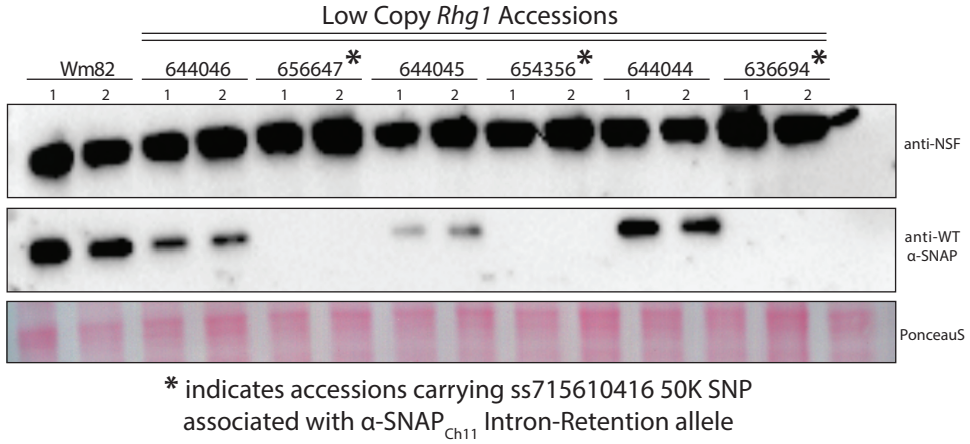


Fig. S10. Low copy *Rhg1* varieties carrying ss715610416 SNP associated with Ch11 α-SNAP-Intron Retention (IR) allele have lower WT α-SNAP abundance. (*A*) Immunoblots for WT α-SNAP and NSF of Wm 82 or of low copy *Rhg1* accessions PI 644046, PI 656647, PI 644045 PI 654356, PI 636694, which are positive (*) or negative for the ss715610416 SNP associated with the *α-SNAP$_{Ch11}$-IR* allele. PonceauS indicates relative total protein abundance per each line.

Soybean Accession	*Rhg1* Copy #	SCN Resistant	*Rhg1* WT α-SNAP	*Rhg1* HC α-SNAP	*Rhg1* LC α-SNAP	Ch11 WT α-SNAP	Ch11 I.R. α-SNAP	Ch07 NSF WT	Ch07 NSF RAN07
Williams 82	1	—	✓	—	—	✓	—	✓	—
PI 548402	3	✓	—	—	✓	—	✓	—	✓
PI 90763	3	✓	—	—	✓	—	✓	—	✓
PI 437654	3	✓	—	—	✓	—	✓	—	✓
PI 89772	3	✓	—	—	✓	—	✓	—	✓
PI 548316	7	✓	✓	✓	—	—	✓	—	✓
PI 88788	9	✓	✓	✓	—	✓	—	—	✓
PI 209332	10	✓	✓	✓	—	✓	—	—	✓

Table S1. *Rhg1* copy number and relevant α-SNAP and NSF alleles present in Wm82 or in the SCN-resistance phenotyping "HG-Type" soybeans. *Rhg1* haplotypes color coded: blue (WT, Single Copy *Rhg1*), red (LC, Low Copy *Rhg1*) or orange (HC, High Copy *Rhg1*). A grey checkmark indicates presence of certain trait or allele and a black minus sign denotes absence. WT is Wild-type allele, I.R. is intron-retention allele of Ch11 α-SNAP (*Glyma.11G234500*) and RAN07 is *Rhg1* associated NSF on Ch07 allele of *Glyma.07G195900*.

Line	*Rhg1* Haplotype	NSF_{Ch07}	NSF_{Ch13}	$\alpha\text{-}SNAP_{Ch11}$
Peking	*Rhg1* $_{LC}$	*Rhg1* Assoc. Allele	WT (Wm82-type)	Intron Retention
90763	*Rhg1* $_{LC}$	*Rhg1* Assoc. Allele	V555I	Intron Retention
437654	*Rhg1* $_{LC}$	*Rhg1* Assoc. Allele	WT (Wm82-type)	Intron Retention
209332	*Rhg1* $_{HC}$	*Rhg1* Assoc. Allele	V555I	WT
89772	*Rhg1* $_{LC}$	*Rhg1* Assoc. Allele	V555I	Intron Retention
548316	*Rhg1* $_{HC}$	*Rhg1* Assoc. Allele	V555I	Intron Retention
Prohio	Susceptible	WT (Wm82-type)	V555I	WT
NE3001	Susceptible	WT (Wm82-type)	Y260F	WT
4J105-34	*Rhg1* $_{HC}$	*Rhg1* Assoc. Allele	V555I, L738F	WT
CL0J095-46	*Rhg1* $_{HC}$	*Rhg1* Assoc. Allele	V555I	WT
IA3023	Susceptible	WT (Wm82-type)	V555I	WT
LD00-3309	*Rhg1* $_{HC}$	*Rhg1* Assoc. Allele	WT (Wm82-type)	WT
LD02-4485	*Rhg1* $_{HC}$	*Rhg1* Assoc. Allele	WT (Wm82-type)	WT
LG05-4292	*Rhg1* $_{HC}$	*Rhg1* Assoc. Allele	WT (Wm82-type)	Intron Retention
LD01-5907	*Rhg1* $_{LC}$	*Rhg1* Assoc. Allele	V555I	Intron Retention
LD02-9050	*Rhg1* $_{HC}$	*Rhg1* Assoc. Allele	V555I	WT
Magellan	Susceptible	WT (Wm82-type)	WT (Wm82-type)	WT
Maverick	*Rhg1* $_{HC}$	*Rhg1* Assoc. Allele	V555I	Intron Retention

Table S2. α-SNAP or NSF alleles identified by whole genome sequencing of HG-Type test lines and *Rhg1*-containing NAM parents. All multi-copy *Rhg1* haplotype lines contained a unique *Glyma.07g195900* NSF_{Ch07} allele (*Rhg1* associated NSF on chromosome 07; NSF_{RAN07}). An $\alpha\text{-}SNAP_{Ch11}$ intron-retention allele was present among some, but not all multi-copy *Rhg1* haplotypes. A *Glyma.13G180100* (NSF_{Ch13}) allele was also detected in some but not all *Rhg1* containing HG-Type and NAM lines, but was also found in some SCN-susceptible varieties.

	ss715597431			ss715597413				ss715597410		
Soybean Line	Glyma.07g195900	Glyma.07g195800	Glyma.07g195700	Glyma.07g195600	Glyma.07g195500	Glyma.07g195400	Glyma.07g195300	Glyma.07g195200	Glyma.07g195100
	NSF (RAN07)	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 548402	R_4Q, N_{21}Y, S_{25}N, ^$_{116}$F, M_{181}I	K_3N, F_{137}S	T_{21}A, K_{23}R, G_{109}C, H_{115}Q V_{345}I, D_{364}N, M_{406}T, Q_{618}K	WT	WT	WT	WT	WT	WT
PI 90763	R_4Q, N_{21}Y, S_{25}N, ^$_{116}$F, M_{181}I	K_3N, L_{42}R, F_{137}S	T_{21}A, K_{23}R, G_{109}C, H_{115}Q V_{345}I, D_{364}N, M_{406}T, Q_{618}K	WT	WT	WT	WT	WT	WT
PI 437654	R_4Q, N_{21}Y, S_{25}N, ^$_{116}$F, M_{181}I	K_3N, F_{137}S	T_{21}A, K_{23}R, G_{109}C, H_{115}Q V_{345}I, D_{364}N, M_{406}T, Q_{618}K	WT	WT	WT	WT	WT	WT
PI 209332	R_4Q, N_{21}Y, S_{25}N, ^$_{116}$F, M_{181}I	K_3N, L_{42}R, F_{137}S	T_{21}A, K_{23}R, V_{345}I, D_{364}N, M_{406}T, Q_{618}K	WT	WT	WT	WT	WT	WT
PI 89772	R_4Q, N_{21}Y, S_{25}N, ^$_{116}$F, M_{181}I	K_3N, F_{137}S	T_{21}A, K_{23}R, G_{109}C, H_{115}Q V_{345}I, D_{364}N, M_{406}T, Q_{618}K	WT	WT	WT	WT	WT	WT
PI 548316	R_4Q, N_{21}Y, S_{25}N, ^$_{116}$F, M_{181}I	K_3N, F_{137}S	T_{21}A, K_{23}R, G_{109}C, H_{115}Q V_{345}I, D_{364}N, M_{406}T, Q_{618}K	WT	WT	WT	WT	WT	WT
Magellan	**WT**	L_{42}R, F_{137}S	D_{364}N, M_{406}T	WT	WT	WT	E_{49}G	D_{60}S, S_{64}P	WT
IA3023	**WT**	L_{42}R, F_{137}S	D_{364}N, M_{406}T, Y_{576}F	WT	WT	WT	E_{49}G	D_{60}A, S_{64}P	WT
LD00-3309	R_4Q, N_{21}Y, S_{25}N, ^$_{116}$F, M_{181}I	K_3N, L_{42}R, F_{137}S	T_{21}A, K_{23}R, G_{109}C, H_{115}Q, V_{345}I, D_{364}N, M_{406}T, G_{516}C, Q_{618}K	WT	WT	WT	WT	WT	WT
CLOJ095-4-6	R_4Q, N_{21}Y, S_{25}N, ^$_{116}$F, M_{181}I	K_3N, L_{42}R, F_{137}S	T_{21}A, K_{23}R, G_{109}C, H_{115}Q V_{345}I, D_{364}N, M_{406}T, Q_{618}K	WT	WT	WT	WT	WT	WT
Maverick	R_4Q, N_{21}Y, S_{25}N, ^$_{116}$F, M_{181}I	K_3N, F_{137}S	T_{21}A, K_{23}R, G_{109}C, H_{115}Q V_{345}I, D_{364}N, M_{406}T, Q_{618}K	WT	WT	WT	WT	WT	WT

Table S3. Amino acid polymorphisms of genes within the chromosome 07 interval co-segregating with *Rhg1*. Polymorphisms are relative to predicted residues of the Williams82 (SCN-susceptible) reference genome. The predicted amino acid sequence of most candidate loci

matches Wm82. Among candidate loci with residue substitutions, only the *NSF RAN07* allele has identical amino acid changes consistent across all *Rhg1*-containing germ- plasm. SCN-susceptible soybean varieties highlighted in green.

Diverse Parent	RR / RR (Ch07/Ch18)	RR / SS (Ch07/Ch18)	SS / RR (Ch07/Ch18)	SS / SS (Ch07/Ch18)
4J105-3-4	41	41	2	31
CL0J095-4-6	35	45	0	37
LD00-3309	38	45	1	27
LD01-5907	32	32	1	42
LD02-4485	37	50	1	28
LD02-9050	43	31	2	34
Maverick	31	34	0	41
LG05-4292	44	41	1	30
Totals	301	319	**8***	270

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 $Rhg1_{HC}$ or $Rhg1_{LC}$ also inherited the NSF_{RAN07} ^116 F and M_{181}I mutations meaning that all 309 RILs that carried the resistance associated *Rhg1* also carried NSF_{RAN07}

Table S4. NSF_{RAN07} co-segregates with *Rhg1* in all *Rhg1*-containing F2:5 offspring derived from *Rhg1*+ X *rhg1*- parental crosses. Segregating lines and 6K SoySNP genotyping were developed and performed in the soybean NAM (nested association mapping) project of Song *et al*., 2017.

Chapter 3: The *rhg1-a* (*Rhg1* low-copy) nematode resistance source harbors a copia-family retrotransposon within the *Rhg1*-encoded α-*SNAP* gene

Bayless, A.M., Zapotocny, R.W., Han, S., Grunwald, D.J., Amundson, K.K. and Bent, A.F. (2019). The rhg1-a (Rhg1 low-copy) nematode resistance source harbors a copia-family retrotransposon within the Rhg1- encoded α-SNAP gene. *Plant Direct*, 3(8), p.e00164.

Author contributions: I contributed by performing RT-PCRs on RAC expression, performed some PCRs to screen for RAC insertion, and performed immunoblots to check non-RAC containing, "LC" soybean accessions for α-SNAP expression.

Chapter 3: The *rhg1-a* (*Rhg1* low-copy) nematode resistance source harbors a copia-family retrotransposon within the *Rhg1*-encoded α-*SNAP* gene

Adam M. Bayless[a]* Ryan W. Zapotocny[a]*, Shaojie Han[a], Derrick J. Grunwald[a], Kaela K. Amundson[a], Andrew F. Bent[a,1].

[a]Department of Plant Pathology, University of Wisconsin - Madison, Madison, WI 53706 USA

*These authors contributed similarly.

[1]To whom correspondence may be addressed: Email: afbent@wisc.edu.

Keywords: Plant disease resistance, retrotransposon, soybean cyst nematode, *Rhg1*

3.1 Abstract

Soybean growers widely use the Resistance to *Heterodera glycines* 1 (*Rhg1*) locus to reduce yield losses caused by soybean cyst nematode (SCN). *Rhg1* is a tandemly repeated four gene block. Two classes of SCN resistance-conferring *Rhg1* haplotypes are recognized: *rhg1-a* ("Peking-type", low copy number, 3 or fewer *Rhg1* repeats) and *rhg1-b* ("PI 88788-type", high copy number, 4 or more *Rhg1* repeats). The *rhg1-a* and *rhg1-b* haplotypes encode α-SNAP (alpha-Soluble NSF Attachment Protein) variants α-SNAP$_{Rhg1}$LC and α-SNAP$_{Rhg1}$HC respectively, with differing atypical C-terminal domains, that contribute to SCN-resistance. Here we report that *rhg1-a* soybean accessions harbor a copia retrotransposon within their *Rhg1* *Glyma.18G022500* (α-SNAP-encoding) gene. We termed this retrotransposon "*RAC*", for *Rhg1* alpha-SNAP copia. Soybean carries multiple *RAC*-like retrotransposon sequences. The *Rhg1 RAC* insertion is in the *Glyma.18G022500* genes of all true *rhg1-a* haplotypes we tested and was not detected in any examined *rhg1-b* or *Rhg1$_{WT}$* (single-copy) soybeans. *RAC* is an intact element residing within intron 1, anti-sense to the *rhg1-a α-SNAP* open reading frame. *RAC* has intrinsic promoter activities, but overt impacts of *RAC* on transgenic α-SNAP$_{Rhg1}$LC mRNA and protein abundance were not detected. From the native *rhg1-a RAC^{+}* genomic context, elevated α-SNAP$_{Rhg1}$LC protein abundance was observed in syncytium cells, as was previously observed for α-SNAP$_{Rhg1}$HC (whose *rhg1-b* does not carry *RAC*). Using a SoySNP50K SNP corresponding with *RAC* presence, just ~42% of USDA accessions bearing previously identified *rhg1-a* SoySNP50K SNP signatures harbor the *RAC* insertion. Subsequent analysis of several of these putative *rhg1-a* accessions lacking *RAC* revealed that none encoded *α-SNAP$_{Rhg1}$LC*, and thus they are not *rhg1-a*. *rhg1-a* haplotypes are of rising interest, with *Rhg4*, for combating SCN populations that

exhibit increased virulence against the widely used *rhg1-b* resistance. The present study reveals another unexpected structural feature of many *Rhg1* loci, and a selectable feature that is predictive of *rhg1-a* haplotypes.

3.2 Introduction

To thrive in their natural environments, organisms must continually sense and respond to changing conditions, including biotic and abiotic stresses. Transposable elements can cause relatively stable variation in numerous plant phenotypes such as flowering time, trichome presence or fruit size (Lisch, 2013). Transposable elements (TEs) may insertionally disrupt genes, or if TE activity is repressed by epigenetic transcriptional silencing, small interfering RNAs and chromatin condensation, this can impact the expression of nearby genes (Sigman and Slotkin, 2016). TEs are also increasingly being identified as modulatory factors during periods of host stress (i.e., heat, pathogen) (Liu et al., 2004; Xiao et al., 2008; Ding et al., 2015). For instance, cis-regulatory motifs within certain TEs can recruit stress-responsive host transcriptional factors, thereby influencing nearby host gene expression and potentially conferring a host benefit (Slotkin and Martienssen, 2007; Woodrow et al., 2010; Matsunaga et al., 2012; McCue and Slotkin, 2012; Cavrak et al., 2014; Makarevitch et al., 2015; Matsunaga et al., 2015; Negi et al., 2016; Galindo-Gonzalez et al., 2017). Some TEs also beneficially modulate the expression of host plant defense or susceptibility genes (Tsuchiya and Eulgem, 2013; Berg et al., 2015). Additionally, numerous studies report TEs lying within or adjacent to putative plant immune genes, however, potential influences on host genes or positive effects are often not apparent (Bhattacharyya et al., 1997; Henk et al., 1999; Wawrzynski et al., 2008).

Glycine max (soybean) is an important food and industrial crop (Schmutz et al., 2010). A major pest afflicting global soybean production is the soybean cyst nematode (SCN, *Heterodera glycines*), which causes yearly U.S. soybean yield losses of over 1 billion USD (Niblack et al., 2006; Mitchum, 2016; T. W. Allen, 2017). SCN is an obligate parasite that invades host roots and induces individual host cells to form a complex syncytium that serves as the SCN feeding site (Niblack et al., 2006; Mitchum, 2016). SCN feeding depletes available host resources and a functional syncytium must be maintained for 2-4 weeks for the nematode to complete its lifecycle. Since the unhatched eggs within cysts can remain viable for many years in the field, SCN is difficult to manage and is primarily controlled by growing naturally resistant soybeans (Niblack et al., 2006). Among known soybean loci contributing to SCN-resistance, the *Rhg1* (Resistance to *Heterodera glycines* 1) locus found on chromosome 18 provides the strongest protection (Concibido et al., 2004). *Rhg1* causes the SCN-induced syncytium to fail a few days after induction, and the soybean PI 88788-type "*rhg1-b*" haplotype is the primary SCN resistance locus used in commercially grown soybeans (Concibido et al., 2004; Niblack et al., 2006; Mitchum, 2016).

Soybean *Rhg1* is an unusual disease resistance locus that consists of a ~31.2 kb unit that is tandemly repeated as many as 10 times (Cook et al., 2012). Within each 31.2 kb *Rhg1* repeat unit are four different *Rhg1*-encoded genes: *Glyma.18G022400*, *Glyma.18G022500*, *Glyma.18G022600*, and *Glyma.18G022700*, none of which have similarity to previously identified resistance genes (Cook et al., 2012; Cook et al., 2014; Lee et al., 2015). Of the three *Rhg1* genes that contribute to SCN-resistance, only *Glyma.18G022500*, an α-SNAP (alpha-Soluble NSF Attachment Protein), has amino acid polymorphisms relative to the wild-type *Rhg1*

gene alleles present in SCN-susceptible soybeans (Cook et al., 2012; Cook et al., 2014; Lee et al., 2015). The mRNA transcript abundance of all three resistance-associated *Rhg1* genes is significantly elevated in SCN-resistant multi-copy *Rhg1* soybeans, relative to SCN-susceptible single-copy *Rhg1* (WT *Rhg1*) soybeans (Cook et al., 2012; Cook et al., 2014).

At least two distinct *Rhg1* genotype classes exist: "low-copy *Rhg1*" (*rhg1-a,* sometimes referred to as $Rhg1_{LC}$, often derived from PI 548402 'Peking'), and "high-copy *Rhg1*" (*rhg1-b*, sometimes referred to as $Rhg1_{HC}$, often derived from PI 88788) (Niblack et al., 2002; Brucker et al., 2005; Cook et al., 2012; Cook et al., 2014; Bayless et al., 2018). These *Rhg1* genotype classes represent two distinct multi-copy *Rhg1* haplotypes that vary most notably by: a) *Rhg1* repeat number (a high or low number of *Rhg1* repeats), and b) encoding distinctive resistance-type α-SNAP proteins with C-terminal polymorphisms at a conserved functional site (Cook et al., 2014; Bayless et al., 2016). *rhg1-a* resistance is bolstered by an unlinked chromosome 8 locus, *Rhg4*, whose presence contributes to full-strength "Peking-type" SCN resistance (Meksem et al., 2001; Liu et al., 2012). *Rhg4* encodes a polymorphic serine hydroxy methyl transferase with altered enzyme kinetics, but the molecular basis of resistance augmentation by *Rhg4* is not yet understood (Liu et al., 2012; Mitchum, 2016). Several *rhg1-b* and *rhg1-a* accessions have been analyzed by whole genome sequencing (WGS) studies, and characteristic single nucleotide polymorphism (SNP) signatures predictive of *rhg1-b* or *rhg1-a* haplotype soybeans have been reported (Cook et al., 2014; Lee et al., 2015; Shi et al., 2015; Kadam et al., 2016; Patil et al., 2019). Additionally, studies by Arelli, Young and others have profiled SCN-resistance among thousands of USDA soybean accessions and noted substantial phenotypic variation (e.g., (Anand, 1984; Hussey et al., 1991; Young, 1995; Diers et al., 1997; Arelli et al., 2000; Vuong et

al., 2015; Klepadlo et al., 2018)). However, the influence of all *Rhg1* haplotype and/or allelic variation factors on SCN resistance expression or plant yield is not yet fully understood.

Several recent studies have deepened our understanding of *Rhg1* molecular function and highlight a central role of the SNARE (Soluble NSF Attachment Protein REceptors)-recycling machinery in SCN-resistance (Matsye et al., 2012; Cook et al., 2014; Bayless et al., 2016; Lakhssassi et al., 2017; Bayless et al., 2018). α-SNAP, and the ATPase NSF (N-ethylmaleimide Sensitive Factor), are conserved eukaryotic housekeeping proteins that form the core SNARE-recycling machinery. They sustain the pool of fusion-competent SNAREs necessary for new membrane fusion events (Sudhof and Rothman, 2009; Zhao et al., 2015). While most animals encode single NSF and α-SNAP genes, soybean is a paleopolyploid that encodes two NSF, four or five α-SNAP and two γ-SNAP (gamma-SNAP) genes, respectively. A C-terminal α-SNAP domain conserved across all plants and animals recruits NSF to SNARE-bundles and stimulates the ATPase activity of NSF that powers SNARE-complex recycling. However, it is this otherwise conserved α-SNAP C-terminal region that is atypical among both *rhg1-b*- and *rhg1-a*-encoded α-SNAP proteins, and accordingly, both *Rhg1* resistance type α-SNAPs are impaired in promoting normal NSF function and instead mediate dosage-dependent cytotoxicity (Sudhof and Rothman, 2009; Cook et al., 2014; Zhao et al., 2015; Bayless et al., 2016). The abundance of the atypical *rhg1-b* α-SNAP$_{Rhg1}$HC protein specifically increases in the SCN feeding site and contributes to *Rhg1*-mediated collapse of the SCN-induced syncytium (Bayless et al., 2016).

At least two additional loci associated with SCN-resistance are also components of the SNARE-recycling machinery (Lakhssassi et al., 2017; Bayless et al., 2018). Recently, a specialized allele of NSF, *NSF$_{RAN07}$* (*Rhg1*-associated NSF on chromosome 07), was shown to be necessary

for the viability of *Rhg1*-containing soybeans (Bayless et al., 2018). Compared to the WT NSF_{Ch07} protein, the NSF_{RAN07} protein more effectively binds to resistance type α-SNAPs and confers better protection against resistance-type α-SNAP-induced cytotoxicity (Bayless et al., 2018). During the *Rhg1*-mediated resistance response, the ratio of *Rhg1* resistance-type to WT α-SNAPs increases and is apparently an important factor underlying resistance (Bayless et al., 2016; Bayless et al., 2018). Two genetic events sharply reduce WT α-SNAP protein abundance in SCN-resistant *rhg1-a* soybeans (Bayless et al., 2018). First, the wild type α-SNAP-encoding block at *Rhg1* on chromosome 18 - a predominant source of total WT α-SNAP proteins in soybean – is absent from all examined *rhg1-a* accessions, thereby diminishing overall WT α-SNAP protein abundance (Cook et al., 2014; Bayless et al., 2018). Secondly, *rhg1-a* lines often carry a null allele of the α-SNAP encoded on chromosome 11 (*Glyma.11G234500*) - the other major source of WT α-SNAP proteins – due to an intronic splice site mutation that causes premature translational termination and loss of protein stability (Matsye et al., 2012; Cook et al., 2014; Lakhssassi et al., 2017; Bayless et al., 2018). Together, the above studies support a paradigm whereby *Rhg1* and associated SCN-resistance loci rewire major components of the soybean SNARE-recycling machinery. Importantly, soybean accessions that carry *rhg1-a* and *Rhg4* can resist many of the virulent SCN populations that partially overcome *rhg1-b* resistance (Brucker et al., 2005; Niblack et al., 2008; Bayless et al., 2016). Therefore, there is considerable interest in understanding and using *rhg1-a*, the subject of the present study, as an alternative to *rhg1-b* in commercial soybean cultivars (Brucker et al., 2005; Liu et al., 2012; Yu et al., 2016).

Presence/absence variation of transposable elements at specific loci is common among different soybean accessions, and tens of thousands of non-reference genome TE insertions

occur between cultivated and wild soybean (Tian et al., 2012). Moreover, high TE densities near genomic regions exhibiting structural polymorphisms such as copy number variation are also reported in soybean (McHale et al., 2012). While examining the *Rhg1* low-copy (*rhg1-a*) haplotype of soybean accession PI 89772, we uncovered an intact copia retrotransposon within all three copies of the *Rhg1*-encoded *α-SNAP* genes. We termed this retrotransposon "*RAC*", for *Rhg1* alpha-SNAP copia). The *RAC* element, which is entirely within the first intron of the *Glyma.18G022500* (α-SNAP) gene, appears to be intact and transcribes anti-sense to the *α-SNAP* ORF. BLAST searches revealed similar copia elements across the soybean genome, suggesting why assemblies of Illumina short-read whole-genome sequences failed to include this sequence within *rhg1-a* assemblies. This *α-SNAP-RAC* insertion was absent from all examined single-copy *Rhg1* (SCN-susceptible) and high-copy *rhg1-b* ($Rhg1_{HC}$) accessions. More than half of the USDA accessions with SoySNP50K SNPs preliminarily indicative of a low-copy *rhg1-a* haplotype did not carry *RAC*, but sub-sampling among those accessions revealed that they do not encode α-$SNAP_{Rhg1}$LC and thus are not *rhg1-a*. The increasingly important *rhg1-a* SCN-resistant soybean breeding lines do harbor this previously unreported retrotransposon within the α-$SNAP_{Rhg1}$LC-encoding gene.

3.3 Results

Multiple *rhg1-a* haplotypes harbor an intronic copia retrotransposon (*RAC*) within the *Rhg1*-encoded *α-SNAP*

The α-SNAPs encoded by the *rhg1-a* and *rhg1-b* loci play a key role in SCN-resistance (Cook et al., 2012; Cook et al., 2014; Bayless et al., 2016; Liu et al., 2017; Bayless et al., 2018).

While entire 31.2 kb *Rhg1* repeats of *rhg1-b* and *Rhg1*$_{WT}$-like haplotypes have been sub-cloned and characterized (Cook et al., 2012), we sought to study the native genomic *rhg1-a* α-SNAP-encoding region to investigate potential regulatory differences between *rhg1-a* and *rhg1-b*. Fig 1A provides a schematic of the 31.2 kb *Rhg1* repeat unit with the four *Rhg1*-encoded genes, while Fig 1B presents a schematic of the previously published *Rhg1* haplotypes (single-, low-, and high-copy) and the respective C-terminal amino acid polymorphisms of their *Rhg1*-encoded α-SNAP proteins (Cook et al., 2012; Cook et al., 2014). Working from previously generated WGS data, we PCR-amplified the native genomic *rhg1-a* α*-SNAP* locus from PI 89772, and unexpectedly, obtained PCR amplicons ~5 kb larger than WGS-based estimates (Fig S1A) (Cook et al., 2014; Liu et al., 2017). Overly large amplicons were also obtained using different PI 89772 genomic DNA templates and/or other PCR cycling conditions, and no *rhg1-a* α*-SNAP* amplicons of expected size were observed (Fig S1A). Sanger DNA sequencing of these unusually sized *rhg1-a* α*-SNAP* amplicons matched WGS predictions until part way into α*-SNAP* intron 1, where a 4.77 kb element was inserted. Immediately following this 4.77 kb insertion, the amplicon sequence again matched WGS predictions (Fig 1C).

An NCBI nucleotide BLAST of the unknown 4.77 kb region returned hits for conserved features of the Ty-1 copia retrotransposon superfamily. Notably, the multi-cistronic open reading frame (ORF) of this copia element was fully intact, and both 5' and 3' LTRs were present (Long Terminal Repeats; LTRs function as transcriptional promoters and terminators, respectively) (Havecker et al., 2004). Subsequently, we named this insert '*RAC*' for *Rhg1* alpha-SNAP copia. Fig 1D shows the *rhg1-a* α*-SNAP-RAC* structure and Fig S2 provides the complete *RAC* nucleotide sequence and highlights the α*-SNAP*$_{Rhg1}$*LC* intron 1 sequences directly flanking

the *RAC*-integration. The *RAC* insertion effectively doubles the pre-spliced α-SNAP$_{Rhg1}$LC mRNA transcript from 4.70 kb to 9.47 kb, yet *RAC* apparently splices out effectively, as all reported cDNAs of mature α-SNAP$_{Rhg1}$LC transcripts do not contain any *RAC* sequences (Cook et al., 2014; Lee et al., 2015; Liu et al., 2017). The *RAC* ORF is uninterrupted and encodes a 1438 residue polyprotein with conserved copia retrotransposon motifs for GAG (group specific antigen) protease, integrase and reverse transcriptase (Peterson-Burch and Voytas, 2002; Havecker et al., 2004; Kanazawa et al., 2009). These conserved RAC polyprotein motifs are highlighted in Fig S3. Intriguingly, *RAC* integrated just 396 bp downstream of the *α-SNAP$_{Rhg1}$LC* start codon, and in an anti-sense orientation (Fig 1D). The intact LTRs and uninterrupted ORF suggest that *RAC* integration could have been a relatively recent event, and that *RAC* may remain functional.

PI 89772 (used above) is one of seven *rhg1-a* and *rhg1-b* soybean accessions used to determine the HG type of SCN populations (Niblack et al., 2002). We subsequently tested for *RAC* insertions within the *Rhg1*-encoded *α-SNAP*(s) genes in the other six soybean accessions used in HG type tests. The *RAC* integration within the PI 89772-encoded *α-SNAP$_{Rhg1}$LC* creates unique 5' and 3' sequence junctions within the *α-SNAP$_{Rhg1}$LC* intron 1, and substantially increases the distance from *α-SNAP* exon 1 to exon 2 from ~400 bp to ~5,000 bp (Fig 1B and C). Therefore, to screen for *α-SNAP-RAC* insertions, we devised PCR assays specific for *RAC-α-SNAP* junctions, or for wild-type (uninterrupted) "WT junctions" separated by the genomic distances from exon 1 to exon 2 that are annotated in the soybean reference genome (Schmutz et al., 2010), as depicted in Fig 1C and 1D. Among all *rhg1-a* haplotype HG type test accessions (PI 90763, PI 89772, PI 437654 and PI 548402(Peking)-derived 'Forrest'), both 5' and 3' *α-SNAP-RAC* junctions were detected (Fig 1D).

RAC is absent from _rhg1-b_ and single-copy _Rhg1$_{WT}$_ accessions

The _α-SNAP-RAC_ junctions were absent from the _rhg1-b_ accessions tested (PI 88788, PI 548316, PI 209332), which instead gave WT junction PCR products (Fig 1D). Because no SNPs exist at the WT junction primer sites across any of the _rhg1-b_ repeats (Cook et al., 2014), absence of 5' _RAC_ junction and 3' _RAC_ junction PCR products for the _rhg1-b_ accessions suggests that those accessions do not carry the _RAC_ copia element in the first intron of their _α-SNAP_ gene. Accession Williams 82 (Wm82, SCN-susceptible, _Rhg1_ single-copy), the source of the soybean reference genome, also gave a product for the WT junction reaction, and no PCR product for a _RAC_ integration within the _Rhg1$_{WT}$_ (single-copy) _α-SNAP_ gene, consistent with the reference genome annotation (Schmutz et al., 2010) (Fig S1B). _RAC_ absence from _rhg1-b_ and WT _Rhg1_ repeats is also consistent with previous studies that sub-cloned and Sanger-sequenced large-insert genomic fragments spanning entire _rhg1-b_ and _Rhg1$_{WT}$_-like repeats and noted no unusual insertions (Cook et al., 2012). Although all seven HG type test soybean lines were previously analyzed via WGS, the _RAC_ insertion was evidently omitted from the four _rhg1-a_ accession assemblies during Illumina short sequence read filtering that excludes repetitive genome elements, and hence from subsequent read mapping and assembly to the Williams 82 reference genome, which lacks the _RAC_ insertion (Cook et al., 2014; Lee et al., 2015). The _RAC_ insertion was apparently missed in other studies due to sequencing of post-splicing _Rhg1 α-SNAP_ cDNAs (Bayless et al., 2016; Liu et al., 2017).

The WT junction PCR experiment also interrogated if _RAC_ is present within each encoded _α-SNAP_ gene of all three _rhg1-a_ repeats. Among all _rhg1-a_ accessions, no WT exon 1 to exon 2 junctions were detected, while uninterrupted WT junction product distances were

present in all *rhg1-b* accessions and Wm82 as noted above (Fig 1E, F, Fig S1B). Hence *RAC* is apparently present within the *α-SNAP* of each *rhg1-a* repeat unit. To independently investigate the same question, previously available Illumina whole genome sequence data were queried (Cook et al., 2014). The read-depth for *RAC* was found to be 3-4 fold greater in *rhg1-a* accessions relative to the read-depth of the flanking DNA regions, or when compared to *RAC* read depth in *rhg1-b* accessions (Fig S4, SI Spreadsheet). Table 1 provides a summary of the *Rhg1* haplotype composition, copy number, resistance-type *α-SNAP* allele, and presence of normal vs. *RAC*-interrupted *α-SNAP* among the HG type test accessions and the Wm82 reference genome. Together, these findings indicate that the HG type test *rhg1-a* accessions contain the *RAC-α-SNAP* introgression, and that their *rhg1-a* repeats are ~36.0 kb, as opposed to ~31.2 kb for *rhg1-b* repeats and $Rhg1_{WT}$. Although *RAC* is integrated in anti-sense orientation and close to α-$SNAP_{Rhg1}LC$ exon 1, *RAC* does not eliminate *rhg1-a* function, because the HG type test accessions are selected owing to their strong SCN-resistance, and, all have previously been shown to express α-$SNAP_{Rhg1}$LC mRNA and protein (Cook et al., 2014; Bayless et al., 2016; Liu et al., 2017; Bayless et al., 2018). *RAC* presence within these *rhg1-a* lines may even be beneficial.

RAC is not identical to other copia elements but the _RAC_-like copia subfamily is common in soybean

Copia retrotransposons frequently attain high copy numbers in plant and animal genomes, therefore, we assessed the abundance of *RAC* and *RAC*-like copia elements in the soybean genome (Du et al., 2010; Du et al., 2010; Zhao and Ma, 2013). SoyTE, the soybean transposon database, has recorded over 32,000 transposable elements, including nearly 5,000

intact retrotransposons (Du et al., 2010). We queried the SoyTE database via Soybase.org using a *RAC* nucleotide sequence BLASTN search, but no intact or high identity hits were returned. However, a similar BLASTN search against the Wm 82 soybean reference genome at Phytozome.org (Goodstein et al., 2012) returned 146 sequences. These 146 hits spanned all 20 *G. max* chromosomes and included several intact elements of high nucleotide identity with *RAC*, as well as numerous short length matches which likely represent fragments of inactive elements (Table S1). We then constructed a nucleotide-based phylogenetic tree of the soybean *RAC* family using just one *RAC*-family element from each soybean chromosome (Chr) (Fig 2A). This analysis used the element from each chromosome that was most similar to *RAC*, as well as a previously reported copia retrotransposon (TGMR) residing near the soybean *Rps1-k* resistance gene (Bhattacharyya et al., 1997) and the highest *RAC*-identity element match from the common bean (*Phaseolus vulgaris*) genome (Fig 2A). The two *RAC*-like elements, from Chr10 and Chr18, had 99.7% and 97.6% nucleotide identity with *RAC*, respectively (Fig 2A, Table S1). Moreover, the Chr10 element retained an intact ORF and both LTRs. The near-perfect nucleotide identity with *RAC* (99.7%; 4464/4477 positions) suggests the Chr10 element as a possible source for the retrotransposition event that created the *RAC* introgression within α-*SNAP*$_{Rhg1}$*LC*. The above-noted WGS read-depth analysis of soybean accessions (Fig S4, SI Spreadsheet) also found that 3 of 10 examined *rhg1-b* accessions gave a read-depth of zero or close to zero for *RAC* (with one mismatch allowed), indicating absence of the *RAC* or a close homolog at the Chr 10 locus in some soybean accessions. Like *RAC*, the Chr18 element was also integrated anti-sense within a host gene *Glyma.18G268000* (a putative leucine-rich repeat receptor kinase). We further noted that the Chr20 *RAC*-like element (82% identical) was

intronically positioned within *Glyma.20G250200* (BAR-domain containing protein) (Table S1). That multiple intact and highly similar *RAC*-like elements are in soybean suggests that this retrotransposon family was recently active.

Subsequent work examined if *RAC*-family elements are present among other plant species. We performed a TBLASTN search of the *RAC*-encoded polyprotein at NCBI and obtained numerous hits against multiple species, including *Arabidopsis*, *Cajanus cajans* (pigeon pea), *Vigna angularis* (adzuki bean), *Phaseolus vulgaris* (common bean)*, Lupinus angustifolius* (blue lupin), *Medicago truncatula*, and clover (*Trifolium subterraneum*). Similar to Fig 2A, a phylogenetic tree using MEGA (muscle alignment) and the *RAC*-family polyprotein sequences of these different plant species was constructed (Kumar et al., 2016) (Fig 2B). Together, these findings demonstrate that the *RAC*-like copia members are not only common in *Glycine max,* but also in other legumes.

RAC is present within only a subclass of the soybean accessions that have the previous SoySNP50K-predicted _rhg1-a_ signature

Although we detected *RAC* in all four *rhg1-a*-containing accessions that are used for HG type determination, we sought to determine if the *RAC* insertion is universal among all *rhg1-a*-containing accessions. Recently, the USDA soybean collection (~20,000 accessions) was genotyped using a 50,000 SNP DNA microarray chip - the SoySNP50K iSelect BeadChip (Song et al., 2015). We searched for and found a SNP on the SoySNP50K chip that detects *RAC*. Using the SoySNP50K browser at Soybase.org (Soybase.org/snps/), we found a SNP (ss715606985, G to A) that in the Wm82 soybean reference genome was assigned to the Chr10 *RAC*-family element

(99.7% nucleotide identity with *RAC*). However, we noted that this ss715606985 SNP is a perfect match to the sequence of *RAC* within α-*SNAP*$_{Rhg1}$*LC* (Fig S5A). Using this SNP marker for *RAC*, we then calculated the ss715606985 SNP prevalence among all USDA accessions, and found that the SNP is rare - only 390 of 19,645 accessions (~2.0%) were putative *RAC*$^{+}$ lines homozygous for the SNP (Fig 3A). The *RAC*-SNP (ss715606985) was then directly tested as a marker for the α-*SNAP-RAC* event using the PCR assays described in Fig 1, which test for α-*SNAP-RAC* junctions and normal α-*SNAP* exon 1-2 distances. We randomly selected several accessions with SNP-signatures of *rhg1-a* that were positive or negative for the *RAC*-SNP, and found that SNP presence correlated perfectly with *RAC*-α-*SNAP* junction detection, while accessions lacking the *RAC*-SNP had normal α-*SNAP* exon 1-2 distances indicative of no inserted DNA (Fig 3B).

We next examined *RAC* presence among all *Glycine max* USDA accessions with the SoySNP50K SNP-signatures of *rhg1-a* or *rhg1-b* haplotypes, as reported by Lee *et al.*(Lee et al., 2015). The multi-SNP SoySNP50K signatures for *rhg1-a* and *rhg1-b* (Lee et al., 2015) are present in 705 and 150 *Glycine max* accessions respectively, out of 19,645 USDA accessions; these SoySNP50K signatures are provided in Fig S5B. We found that 42% (299 of 705) of accessions with *rhg1-a* SNP-signatures and 0% (0 of 150) of accessions with *rhg1-b* SNP-signatures carry the *RAC* ss715606985 SNP (Fig 3C). That the *RAC*-SNP was absent from all *rhg1-b* signature accessions is consistent with the PCR screens of Fig 1E and F, which indicated that no *rhg1-b* HG type test accession contained *RAC*-α-*SNAP* junctions. A flow-chart is available as Supplementary Fig S6 that summarizes the above findings and additional work presented below.

Because only 299 of the 390 accessions with the *RAC*-SNP had a perfect-match *rhg1-a* SoySNP50K SNP signature (Fig 3C), we investigated the *Rhg1* SNP signature of the remainder. To avoid false positives, the *rhg1-a* SNP signature of (Lee et al., 2015) uses 14 SNPs that extend to 11 kb and 54 kb beyond the edges of the ~30 kb *Rhg1* repeat. Relaxing the stringency that required 14 perfect matches, we found that 83 of the remaining 91 *RAC* $^{+}$ accessions carry a perfect match with the four *rhg1-a* SNP markers that map within the *Rhg1* repeat or within <5 kb of the edge of the *Rhg1* repeat. 86 of 91 have only a single reliably called SNP that varies from the *rhg1-a* consensus (SI Spreadsheet). Together, the combined findings indicate that the *α-SNAP-RAC* integration is only present within *rhg1-a* haplotypes, and that *RAC* retrotransposition may have occurred within a subset of the *Rhg1*$^{+}$ population after *Rhg1* divergence into the distinctive high and low copy haplotype classes.

The *RAC*-SNP allows more accurate prediction of *rhg1-a* presence

The above finding that a few hundred of the 705 USDA accessions with the previously identified *rhg1-a* SNP signature apparently do not contain *RAC* was surprising, given that all four of the *rhg1-a* HG type test accessions do contain *RAC-α-SNAP* junctions. However, it was possible that these non-*RAC* containing accessions, despite a consensus SNP-signature predicting an *rhg1-a*-haplotype, might not truly carry an *rhg1-a* resistance haplotype. *rhg1-a* and *rhg1-b* repeats encode distinct *Rhg1 α-SNAP* alleles, thus, we cloned and sequenced the genomic *Rhg1 α-SNAP* regions from several non-*RAC rhg1-a* SNP-signature accessions and detected coding sequences for either *rhg1-b* (*α-SNAP*$_{Rhg1}$*HC*) or *Rhg1*$_{WT}$ (*α-SNAP*$_{Rhg1}$*WT*) alleles (Fig S5C). None of these accessions encoded *α-SNAP*$_{Rhg1}$*LC* and thus, they were not *rhg1-a* (Fig

S5C). These findings indicate that the consensus *rhg1-a* SNP-signature, while useful, is not a perfect predictor of accessions carrying actual *rhg1-a* resistance haplotypes. Rather, combined presence of the ss715606985 SNP for *RAC* and a near-consensus *rhg1-a* signature is a more accurate predictor of accessions that truly carry *rhg1-a* resistance. Additionally, these data suggest that accessions carrying *rhg1-b* resistance haplotypes can share the SNP signatures of *rhg1-a* accessions. We again refer readers to the flow-chart (Supplementary Fig S6) that summarizes these and other findings.

RAC presence correlates with a stronger SCN-resistance profile and the co-presence of other loci that augment _rhg1-a_ resistance

rhg1-a (*Rhg1* low copy) loci encode unique α-*SNAP*$_{Rhg1}$*LC* alleles, however, robust *rhg1-a* resistance requires the co-presence of *Rhg4*, and the α-*SNAP Ch11-IR* allele bolsters *rhg1-a* resistance further (Liu et al., 2012; Lakhssassi et al., 2017; Bayless et al., 2018; Patil et al., 2019). We sought to compare the SCN resistance profiles of the *rhg1-a* signature accessions with *RAC* to those without *RAC* (which are not true *rhg1-a*), to assess how the two groups match what is known about *rhg1-a* resistance. Previously, Arelli, Young, and others obtained SCN resistance phenotype data, across multiple trials and with different SCN populations, for at least 573 different USDA accessions that are now known to carry the SoySNP50K signatures suggestive of *rhg1-a* (Anand, 1984; Hussey et al., 1991; Young, 1995; Diers et al., 1997; Arelli et al., 2000; Lee et al., 2015). We used these available SCN resistance data from the USDA GRIN database to compare the resistance profiles of *rhg1-a*-signature accessions which did or did not have the ss715606985 (*RAC*) SNP signature. Any accession that scored as “R” (resistant) in any single SCN

trial was placed into the broad category “R”. Likewise, any accession that scored “MR” (moderately resistant) or “MS” (moderately susceptible) in any trial, with no higher resistance scores in other trials, was placed into those respective categories. Only accessions that scored susceptible (“S”) across all trials were placed into the “S” category. Consistent with previous reports that *rhg1-a* accessions possess broad and robust resistance(Concibido et al., 2004; Vuong et al., 2015; Kadam et al., 2016), 91% (51/56) of the accessions in the “R” group were positive for the ss715606985^{+} *RAC* SNP (Fig 4A). The frequency of *RAC* presence was substantially lower among the more susceptible phenotypic classes (Fig 4A) while the majority of the ss715606985^{-} (no *RAC*) accessions scored either “S” or “MS”. As was noted above, none of the non-*RAC* accessions that we examined had *rhg1-a* type resistance (Fig 3B, 3C, Fig S5C). The phenotype scores and relevant SNP markers for all 573 of these SCN-phenotyped *rhg1-a* SNP signature accessions are provided as a spreadsheet in the SI Data.

In the above analysis (Fig 4A), some of the *RAC*$^{+}$ (ss715606985^{+}) accessions, which are *rhg1-a*, had scored as "S" or "MS". This seemed likely to be because they lack a resistance-conferring allele at *Rhg4* and/or the resistance-enhancing allele of the Chr 11-encoded α-*SNAP* (α-*SNAP*$_{Ch11}$-*IR*) (Liu et al., 2012; Lakhssassi et al., 2017; Bayless et al., 2018). Accordingly, we investigated if the SCN resistance phenotype scores also correlated with co-presence of those loci. None of the SoySNP50K markers resides within the *Rhg4* gene but we noted that the two SoySNP50K SNPs that most closely flank the *Rhg4* locus are rare among USDA accessions (Fig 4B; ss715602757, ss715602764), and one or both of these SNPs is present in the *Rhg4*-containing HG type test lines. We also used a SNP, ss715610416, previously associated with the Chr11 α-*SNAP* intron-retention allele (α-*SNAP*$_{Ch11}$-*IR*) (Bayless et al., 2018). Among the 51 *RAC*$^{+}$

accessions with an SCN resistance score of "R", we found that SNPs associated with both α-*SNAP*$_{Ch11}$*-IR* and *Rhg4* were enriched ~10-fold, as compared to the entire USDA collection (Fig 4 B,C). Additionally, among the 705 USDA accessions with SoySNP50K signatures predictive of *rhg1-a*, we found that the *Rhg4* and *α-SNAP*$_{Ch11}$*-IR* SNPs were enriched among the *RAC*$^{+}$ accessions relative to the *RAC*$^{-}$ accessions (Fig 4D). Thus, the heightened SCN resistance of the *RAC*-positive (ss715606985^{+}) *rhg1-a*-signature accessions is consistent with previous reports, and as expected, the resistance is associated with the co-presence of additional loci like *Rhg4* and α-*SNAP*$_{Ch11}$*-IR* (Vuong et al., 2015; Kadam et al., 2016; Lakhssassi et al., 2017; Bayless et al., 2018; Patil et al., 2019).

The *rhg1-a RAC* element has intrinsic transcriptional activity

While the *RAC*-SNP apparently identifies true *rhg1-a* resistance sources, possible impacts of the *RAC* element itself on α-*SNAP*$_{Rhg1}$*LC* expression remained to be explored. Typically, eukaryotic cells silence transposable elements using small RNA-directed DNA methylation pathways, and this can also silence adjacent genes (McCue et al., 2012; Kim and Zilberman, 2014). Since the α-SNAP$_{Rhg1}$LC mRNA transcript and protein are readily detected in *RAC*-containing soybeans, *RAC* does not, at least constitutively, eliminate α-*SNAP*$_{Rhg1}$*LC* expression (Cook et al., 2014; Bayless et al., 2016; Liu et al., 2017; Bayless et al., 2018). Nonetheless, we examined DNA methylation at the *rhg1-a* α-*SNAP-RAC* junction, as well as transcriptional activity of the *rhg1-a RAC* element. The restriction enzyme McrBC cleaves only methylated DNA, so potentially methylated DNA regions may be assessed via McrBC digestion and subsequent attempted PCR across areas of interest. After genomic DNAs from 'Forrest' and

'Peking' (PI 548402) were treated with McrBC, both the 5' and 3' borders of the α-*SNAP-RAC* did not PCR amplify relative to the mock treated controls, indicating the presence of methylated cytosines at the α-*SNAP-RAC* junctions (Fig 5A).

Because *RAC* has both LTRs and an intact ORF, we tested for transcription of *RAC* in the *rhg1-a* soybean genotype 'Forrest' as compared to 'Fayette' (*rhg1-b*) and Wm82 ($Rhg1_{WT}$). Fayette and Wm82 do not carry the *Rhg1* α-*SNAP-RAC* but do carry other *RAC*-like copia elements that match the qPCR primers used (diagrammed in Fig 5C). qPCR analysis of cDNAs from root or leaf tissues indicated that mRNA transcripts from *RAC* or *RAC*-like sequences were ~200-fold higher in 'Forrest' than in Wm82 or 'Fayette' (*rhg1-b*) (Fig 5B). This suggested but did not firmly demonstrate that the *Rhg1*-embedded *RAC* is the primary source of the detected transcript, because *RAC* has high nucleotide identity with other *RAC*-like elements (Fig 2A) whose activity may also vary between accessions.

We conducted additional tests for transcription of α-*SNAP-RAC* by transforming Wm82 roots with a ~15 kb cloned segment of native α-*SNAP-RAC* genomic DNA (including the upstream *Glyma.18G024400 Rhg1* gene, which shares the same bidirectional promoter; depicted in Fig 5C). Importantly, we engineered this otherwise native α-*SNAP-RAC* cassette with a unique nucleotide tag to distinguish between transgene derived transcripts vs. transcripts from other *RAC*-family elements in the genome (Fig 5C). Low abundance of *RAC* transcripts in Wm82 roots relative to 'Forrest' roots had been documented (Fig 5B), so RT-PCR of Wm82 readily visualized *RAC*-specific transcript expression from the transgenically introduced construct. In control roots, sharp contrasts in *RAC* expression were again observed between 'Forrest' roots and Wm82 roots transformed with empty vector (Fig 5D). But Wm82 roots

transformed with the uniquely tagged *α-SNAP-RAC* transgene had substantially elevated *RAC* transcripts compared to isogenic Wm82 controls, as indicated by a primer pair that amplifies all *RAC* sequences (native or tagged), and by a primer pair that amplifies only the uniquely tagged *α-SNAP-RAC* transcript (Fig 5D). Controls using template samples prepared without reverse transcriptase verified successful DNAase treatment of cDNA preparations (Fig 5D). We further tested the activity of the *RAC* promoter itself by constructing a native 4.77 kb *RAC* element cassette divorced from the flanking *Rhg1* DNA (Fig 5E), which we then transformed into Wm82. Similar to Fig 5D, the native 4.77 kb *RAC* transgene substantially increased *RAC* transcript abundance in Wm82 roots, relative to empty vector controls (Fig 5F). Together, these findings demonstrate that presence of the *rhg1-a* locus *RAC* can substantially elevate *RAC* mRNA transcripts, and that *RAC* itself possesses intrinsic promoter activity. The findings suggest that the high *RAC* transcript abundance observed in 'Forrest' (*rhg1-a*), but not 'Fayette' (*rhg1-b*) or Wm 82 (single-copy *Rhg1*), is likely to be derived from the *rhg1-a* locus *RAC* insertion. These findings also support the possibility that *RAC* may retain the potential to promote transposition.

α-SNAP$_{Rhg1}$LC protein is expressed despite *RAC* presence

TEs can influence the expression of host genes. Because *RAC* is present in *rhg1-a* accessions previously chosen for use in agricultural breeding due to their strong SCN-resistance, *RAC* presence may benefit *rhg1-a*-containing soybeans. In light of *RAC's* anti-sense orientation and close proximity to the *Rhg1 α-SNAP$_{Rhg1}$LC* promoter, we sought to examine if *RAC* influences α-SNAP$_{Rhg1}$LC protein expression. We were not able to compare expression of α-SNAP$_{Rhg1}$LC between native *rhg1-a* loci that do or do not contain *RAC*, because no *rhg1-a*

accessions without *RAC* have been identified and deleting *RAC* from all *Rhg1* repeats of a *rhg1-a* accession would not be trivial. Therefore, we left intact or removed the 4.77 kb *RAC* insertion from the native *α-SNAP-RAC* construct used for Fig 5 (Fig 6A) and then examined α-SNAP$_{Rhg1}$LC protein abundance in transgenic Wm82 roots carrying the respective constructs. Immunoblotting was conducted using previously described α-SNAP$_{Rhg1}$LC-specific and WT α-SNAP-specific antibodies (Bayless et al., 2016). The results of a representative experiment are shown in Fig 6B. Across multiple experiments containing independently transformed roots, the constitutive expression of α-SNAP$_{Rhg1}$LC protein was highly variable, regardless of *RAC* presence/absence. However, with respect to constitutive expression of α-SNAP$_{Rhg1}$LC protein we observed no requirement for *RAC* nor any obvious detrimental impact of *RAC* (Fig 6B).

Previously, we reported that native α-SNAP$_{Rhg1}$LC mRNA transcripts include an alternative splice product (Cook et al., 2014; Bayless et al., 2016). Because TEs can influence host mRNA splicing (Krom et al., 2008), we also examined how *RAC* influenced splicing of the known α-SNAP$_{Rhg1}$LC alternative transcript. As above, we generated transgenic roots of Wm82 containing either a *RAC*$^{+}$ *α-SNAP*$_{Rhg1}$*LC* native genomic segment or a version with *RAC* precisely deleted (Fig 6A). We then generated cDNAs and performed RT-PCR with primer sets specific for either the full length or shorter splice isoform. As shown by agarose gel electrophoresis, *RAC* presence was not required for alternate splicing of this α-SNAP$_{Rhg1}$LC isoform (Fig 6B).

α-SNAP$_{Rhg1}$LC hyperaccumulates at SCN infection sites

We also examined infection-associated α-SNAP$_{Rhg1}$LC protein expression in non-transgenic soybean roots that carry the native *rhg1-a* locus. We previously reported that

during *rhg1-b*-mediated SCN-resistance, α-SNAP$_{Rhg1}$HC abundance is elevated ~12-fold within syncytial cells (SCN feeding sites) relative to adjacent non-syncytial cells (Bayless et al., 2016). To test whether the *RAC*-containing *rhg1-a* follows a similar expression pattern during the resistance response, the present study examined α-SNAP$_{Rhg1}$LC abundance at SCN infection sites in soybean variety 'Forrest' using SDS-PAGE and immunoblots. Using the aforementioned α-SNAP$_{Rhg1}$LC-specific antibody, we detected increased α-SNAP$_{Rhg1}$LC accumulation within tissues enriched for SCN feeding sites, while expression was barely detectable in mock-treated roots (Fig 7A). As previously reported, NSF proteins were also increased in SCN-infested roots, albeit less prominently (Fig 7A) (Bayless et al., 2016). Thus, even in *RAC* presence, *rhg1-a* haplotypes drive an expression pattern of α-SNAP$_{Rhg1}$LC similar to that observed for *rhg1-b* and α-SNAP$_{Rhg1}$HC.

To more precisely locate the α-SNAP$_{Rhg1}$LC increases, SCN-infested root sections were imaged using transmission electron microscopy and immunogold labelling of bound α-SNAP$_{Rhg1}$LC-specific antibody. Syncytium-specific accumulation of α-SNAP$_{Rhg1}$LC protein was observed and quantified in root sections taken 7 days after SCN inoculation (Fig 7B, C). The average increase of immunogold particles per equal area of adjacent non-syncytial root cells (cells still carrying a large central vacuole) was ~25-fold (Fig 7C). In control experiments, EM sections from mock-inoculated roots (no SCN) exhibited no immunogold signal above background (Fig S7A). Similarly, no immunogold signal above background was observed when secondary antibody and all other reagents were used but the primary antibody was omitted (Fig S7B). The specificity of the antibody for α-SNAP$_{Rhg1}$LC protein was previously demonstrated (signal for recombinant α-SNAP$_{Rhg1}$LC protein or total protein from roots with *rhg1-a*, no signal

for α-SNAP$_{Rhg1}$HC protein or total protein from roots with *rhg1-b* or *Rhg1$_{WT}$*) [30]. The above results, similar to the previously observed ~12-fold increase reported for the α-SNAP$_{Rhg1}$HC in syncytia from *rhg1-b* roots, indicate that α-SNAP$_{Rhg1}$LC protein abundance is also elevated within syncytia upon SCN infection [30]. Collectively, these findings demonstrate that while a potentially active retrotransposon (*RAC*) has integrated within the important *rhg1-a α-SNAP$_{Rhg1}$LC* resistance gene, and its presence correlates with *rhg1-a* haplotypes preferred for SCN resistance breeding, no negative impacts of *RAC* on α-SNAP$_{Rhg1}$LC mRNA or protein expression were detected.

3.4 Discussion

Rhg1 is the principal SCN resistance locus in commercially grown soybeans. The increasing occurrence of SCN populations that at least partially overcome the overwhelmingly utilized "PI 88788-type" *rhg1-b* resistance source is an important concern for soybean breeders and growers (McCarville et al., 2017), (www.thescncoalition.com). Alternating use of different *Rhg1* haplotypes should help bolster and preserve resistance against these virulent SCN populations (Brucker et al., 2005; Niblack et al., 2008)(www.thescncoalition.com). In this study we report that the other *Rhg1* haplotype available for SCN control, *rhg1-a* (also known as "Peking-type" *Rhg1*), carries a distinct genetic structure. *rhg1-a* unexpectedly contains an intact and transcriptionally active retrotransposon within an intron of the key *Rhg1 α-SNAP* resistance gene in each repeat. The "Hartwig-type" SCN resistance from PI 437654 also carries the *rhg1-a* haplotype, and also carries the *RAC* retrotransposon within the *Rhg1 α-SNAP* genes.

Transposons have been coopted for the service of defense responses in both plants and animals (Tsuchiya and Eulgem, 2013; Huang et al., 2016). V(D)J recombination, which underlies the remarkable diversity of vertebrate adaptive immunity, apparently derives from a domesticated RAG-family transposase (Huang et al., 2016). *RAC* has inherent transcriptional activity and is positioned anti-sense within the first intron of the *rhg1-a* α*-SNAP* gene. It is unclear if *RAC* impacts *rhg1-a* function (discussed below). However, this study revealed the utility of *RAC* and the *RAC*-associated ss715606985 (G to A) SNP in more accurately identifying SCN resistance-conferring *rhg1-a* germplasm. Among the 19,645 USDA soybean accessions genotyped using the SoySNP50K iSelect BeadChip (Song et al., 2015), a few hundred accessions with a *rhg1-a*-type SNP signature apparently do not actually carry a *rhg1-a* locus. All of the *rhg1-a*-signature soybeans we examined that do not carry *RAC* encoded *rhg1-b* or *Rhg1*$_{WT}$ α*-SNAP* alleles. Conversely, all examined *rhg1-a*-signature soybeans with *RAC* carried the *rhg1-a* α*-SNAP* allele.

Active retrotransposon families are abundant in soybean (Wawrzynski et al., 2008). *RAC* has similarities to a copia element near a *Phytophthora sojae* resistance locus identified by Bhattacharyya *et al.*, however, SoyTE database searches returned no highly similar *RAC*-family TEs (Bhattacharyya et al., 1997; Du et al., 2010). Although whole-genome sequencing studies previously examined the HG type test soybean accessions, the *RAC* insertion within α*-SNAP*$_{Rhg1}$*LC* was apparently omitted during the filtering steps of DNA sequence read mapping and assembly (Cook et al., 2014; Liu et al., 2017). Our findings revealed multiple *RAC*-family elements in soybean, and this abundance of *RAC*-family elements likely led to *RAC* omission from previous *rhg1-a* sequence assemblies. It is intriguing that *RAC*, at least from PI 89772, has

inherent transcriptional activity and an intact ORF encoding conserved functional motifs – features of an autonomous element. Additionally, *RAC*'s near perfect identity with the Chr10 element supports that *RAC* family retrotransposons were recently active in soybean.

Host silencing of transposable elements can establish *cis*-regulatory networks where the expression of nearby host genes may also be impacted (Lisch and Bennetzen, 2011; McCue et al., 2012; McCue and Slotkin, 2012). In some cases, biotic stresses can influence transposon methylation, and subsequently, alter the expression of host genes near transposons (Dowen et al., 2012). We noted DNA methylation at the *α-SNAP-RAC* junctions. However, we also found evidence that *RAC* is transcriptionally active. In addition, the *Rhg1 α-SNAP* gene that contains *RAC* successfully expresses the α-SNAP$_{Rhg1}$LC protein during SCN-resistance similarly to that observed for α-SNAP$_{Rhg1}$HC. Future analyses of small RNAs may provide evidence of differential silencing of *RAC* or *α-SNAP$_{Rhg1}$LC*. Cyst nematode infection of *Arabidopsis* has been reported to trigger the hypomethyation and activation of certain transposable elements, and moreover, many of these transposable elements reside near host genes whose expression is altered during syncytium establishment (Hewezi et al., 2017; Piya et al., 2017). Thus, it remains an intriguing hypothesis that *RAC* may influence the epigenetic landscape of *α-SNAP$_{Rhg1}$LC* or the overall ~36 kb *rhg1-a* repeat during infection, particular stresses, developmental stages or in specific tissues. Moreover, small RNAs deriving from the other *RAC*-like elements in the soybean genome could modulate *α-SNAP$_{Rhg1}$LC* expression *in trans* (Slotkin and Martienssen, 2007; McCue et al., 2012; McCue et al., 2013). Our BLAST searches revealed at least two other *RAC*-like elements positioned intronically or adjacent to putative host defense and/or

developmental genes. Future studies may interrogate if *RAC*, and/or other endogenous retrotransposons, impact the regulation of host defense gene networks in soybean.

The currently available picture of *Rhg1* haplotype evolution is incomplete and has been dominated by study of lines that are the product of ongoing selection for the most effective SCN-resistance (*e.g.,* modern 10-copy *rhg1-b* and 3-copy *rhg1-a* haplotypes). The finding of *RAC* in all copies of the *rhg1-a* repeat, and in all confirmed *rhg1-a* haplotypes that were tested to date, suggests but does not confirm that *RAC* plays an adaptive role in those haplotypes. The sequence of the *Rhg1* repeat junction is identical between *rhg1-b* and *rhg1-a* haplotypes, as are many SNPs not present in the Williams 82 soybean reference genome, providing evidence of the shared evolutionary origin of *rhg1-b* and *rhg1-a* (Cook et al., 2012). The finding to date of *RAC* only in *rhg1-a* haplotypes suggests that this retroelement probably inserted in *Rhg1* after the divergence of *rhg1-b* and *rhg1-a*. However, it is also possible that the *RAC* retroelement was ancestrally present but then purged from the progenitors of current *rhg1-b* accessions. A correlation has been demonstrated between *Rhg1* copy number and SCN-resistance, and, *rhg1-a* in the absence of *Rhg4* confers only partial SCN resistance (Liu et al., 2012; Cook et al., 2014; Lee et al., 2016; Yu et al., 2016; Kandoth et al., 2017). Yet there are no known instances of *rhg1-a* accessions with an *Rhg1* copy number above three. The *RAC* may allow increased and/or more tightly regulated expression of *rhg1-a*. Alternatively, it is possible that absence of *RAC* is advantageous in allowing increased copy number of *Rhg1*, but that too is only a hypothesis, raised by the present work and in need of future testing. Additional questions about *Rhg1* locus evolution remain that have functional implications for the efficacy of SCN resistance. For example, we know of no *rhg1-a* haplotypes that carry an α-SNAP$_{Rhg1}$WT-encoding *Rhg1* repeat,

which all examined *rhg1-b* haplotypes do contain. Might *RAC* acquisition have influenced this absence of the WT *Rhg1* repeat? Do any α-SNAP$_{Rhg1}$LC-expressing accessions exist that do not carry the *RAC* integration? The USDA soybean collection contains numerous accessions that are positive for *NSF$_{RAN07}$* but which carry *Rhg1* copy numbers below 3 or 10, or have slight deviations from consensus *rhg1-a* or *rhg1-b* SNP signatures. Intensive study of these accessions may shed further light on *Rhg1* haplotype evolution, and moreover, may facilitate the discovery of new and agriculturally useful *Rhg1* alleles.

The finding that popular *rhg1-a* breeding sources contain an intact retrotransposon within *α-SNAP$_{Rhg1}$LC* was surprising, given that these accessions have previously been sequenced multiple times by different groups (Cook et al., 2014; Liu et al., 2017; Patil et al., 2019). The *RAC*-SNP is rare among all USDA accessions, and approximately 700 of 19,645 USDA accessions carry a SoySNP50K signature predictive of an *rhg1-a* haplotype. However, ~400 of these putative *rhg1-a* accessions do not carry the *RAC*-SNP and all of the non-*RAC* putative *rhg1-a* accessions that we sampled did not encode *α-SNAP$_{Rhg1}$LC*, indicating that they are not true *rhg1-a*. Accordingly, most of these non-*RAC* accessions scored phenotypically as SCN-susceptible. Taken together with our PCR assays showing perfect correlation of the *RAC*-SNP with *RAC* presence, our findings indicate that the *RAC*-SNP successfully identifies true *rhg1-a* loci (*i.e.*, those that encode the α-SNAP$_{Rhg1}$LC protein) which, in combination with *Rhg4* and other loci, confers strong SCN resistance.

Correlation of SNP data with previously published SCN resistance phenotype data indicated that the vast majority of "R" scoring *rhg1-a*-signature accessions were *RAC*$^{+}$. Many of the 705 accessions postulated (using the earlier SoySNP50K SNP signature) to be lines that carry

rhg1-a turned out to carry *rhg1-b*, which would explain their resistance to SCN. The large majority of the subset that are not positive for *RAC* were scored as SCN-susceptible or moderately susceptible. Some *RAC*$^{+}$ lines also were scored as SCN-susceptible or moderately susceptible, but most of these are apparently due to the absence of a resistance-associated *Rhg4* and/or the Chr 11 α-*SNAP* intron-retention allele (α-*SNAP*$_{Ch11}$-*IR*), consistent with the established contributions of those loci to SCN resistance (Meksem et al., 2001; Yu et al., 2016; Kandoth et al., 2017). Among the "R" scoring *RAC*$^{+}$ accessions, SNPs genetically linked to *Rhg4* and the Chr 11 α-*SNAP* intron-retention allele were substantially elevated.

Potential modulation of α-*SNAP*$_{Rhg1}$*LC* expression by *RAC*, either during the SCN-resistance response or in certain developmental or stress situations, could benefit *rhg1-a*-containing soybeans that have significantly depleted WT α-SNAP proteins. α-SNAPs, together with NSF, carry out essential eukaryotic housekeeping functions by maintaining SNARE proteins for vesicle trafficking. Notably, among true *rhg1-a* accessions, the abundance of wild-type (WT) α-SNAP proteins is sharply diminished, as compared to *rhg1-b*- or SCN-susceptible soybeans (Bayless et al., 2018). Moreover, α-SNAP$_{Rhg1}$LC protein was shown to be cytotoxic in *Nicotiana benthamiana* while WT α-SNAP co-expression alleviated this toxicity (Bayless et al., 2016). The more recent discovery of the *NSF*$_{RAN07}$ allele as a requisite for the viability of *Rhg1* soybeans further underscores the necessity of the SNARE-recycling machinery for overall plant health (Bayless et al., 2018). The intronic copia element within the *Arabidopsis RPP7* (*Recognition of Peronospora Parasitica 7*) gene serves as an example of a retroelement with immunomodulatory function, having been shown to modulate *RPP7* transcript splicing and expression (Tsuchiya and Eulgem, 2013). However, we did not detect any influence of *RAC* on

constitutive α-SNAP$_{Rhg1}$LC protein expression from transgenes delivered to roots. We also did not detect a functional influence of *RAC* in its native *rhg1-a* haplotype context, insofar as that α-SNAP$_{Rhg1}$LC protein abundance is successfully elevated in syncytia similarly to the reported syncytium elevation of the α-SNAP$_{Rhg1}$HC protein of *rhg1-b* haplotypes, which do not carry *RAC* (Bayless et al., 2016). Hence although we did not find any true *rhg1-a* soybeans without *RAC* integrations, it remains possible that *RAC* integration was a neutral event that confers no host advantage or disadvantage.

SCN causes the most yield loss of any disease for U.S. soybean farmers and *rhg1-a* offers a potential solution to SCN populations that overcome commonly used *rhg1-b* resistance sources (Brucker et al., 2005; Niblack et al., 2008; T. W. Allen, 2017). Findings continue to emerge that further characterize different sources of SCN resistance, including exciting new findings regarding copy number variation at *Rhg4* (Patil et al., 2019). An attractive overall hypothesis for future study of *RAC* is that, in the presence of SCN or other stresses, *RAC* provides an additional regulatory layer to optimize the SCN resistance response mediated by *rhg1-a* and *Rhg4*, and/or promotes plant health in the absence of SCN. By revealing the existence of *RAC* within the important *rhg1-a* haplotype, the present study provides a marker for finding such soybeans, and expands our knowledge regarding the genetic structure and divergence of the agriculturally valuable *Rhg1* source of SCN resistance.

3.5 Materials and Methods

Transgenic soybean hairy root generation and root culturing

Transgenic soybean roots were produced using *Agrobacterium rhizogenes* strain "ARqua1" (Quandt et al., 1993) and the previously described binary vector pSM101, as in (Cook et al., 2012). Transgenic roots were sub-cultured in the dark at room temperature on hairy root medium as in (Cook et al., 2012).

DNA extraction

Soybean genomic DNAs were extracted from expanding trifoliates or root tissues of the respective soybean accessions using standard CTAB methods similar to (Cook et al., 2012).

Amplification and detection of *RAC* (*Rhg1* alpha-SNAP copia)

For initial amplification and subcloning of native *α-SNAP-RAC*, approximately 100 ng of CTAB-extracted gDNA from PI 89772 (*rhg1-a*) was PCR amplified for 35 cycles using HiFi polymerase (KAPA Biosystems, Wilmington, MA). Primer annealing was at ~70°C for 30 seconds and extension was at 72°C for 5 minutes. The resulting *α-SNAP-RAC* amplicon from PI 89772 was separated by agarose gel electrophoresis, gel extracted using a Zymoclean Large Fragment DNA Recovery Kit (Zymo Research, Irvine CA) and TA overhang cloned into a pTopo xL vector using the Topo xL PCR Cloning Kit (Life Technologies Corp., Carlsbad CA), per manufacturer's recommendations. For PCR detection of *α-SNAP-RAC* junctions or WT exon distances, ~25 ng of CTAB-extracted genomic DNA from each respective accession was amplified using GoTAQ Green (New England Biolabs, Ipswich MA) for 32 cycles, separated on a 0.8% agarose gel and visualized.

Phylogenetic tree construction

For the *RAC*-like nucleotide tree, evolutionary analyses were conducted in MEGA7 (Kumar et al., 2016) and evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura-Nei model (Tamura et al., 2004). The tree with the highest log likelihood (-47751.11) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The analysis involved 24 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 4157 positions in the final dataset.

For the *RAC*-polyprotein tree, evolutionary analyses were conducted in MEGA7 (Kumar et al., 2016) and the evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model (Jones et al., 1992). The tree with the highest log likelihood (-12400.87) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 14 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 644 positions in the final dataset.

Read Depth Analysis of *RAC*

Using previously published whole-genome-sequencing data (Cook et al., 2014), read depth was computed using the depth program of SAMtools (Li et al., 2009). Depth was averaged in 250 bp intervals on Chromosome 10 from bp 40,650,000 - 40,690,000 (includes flanking regions of the 99.7% identity *RAC*-like element). The copy number of the Chromosome 10 *RAC*-like element was then calculated as the ratio of the read coverage per 250 bp from bp 40,672,000-40,675,750, divided by the average read coverage for the flanking regions between bp 40,650,000 - 40,690,000. Sequencing coverage was visualized using ggplot (Wickham, 2009) within RStudio (RStudio Team, 2015).

Methylation analysis

McrBC methylation studies were performed similarly to (Cook et al., 2014). Control McrBC reactions contained equivalent amounts of gDNA in reaction buffer, but had no added McrBC enzyme. McrBC digestion was performed at 37ºC for 90 minutes, followed by a 20 minute heat inactivation at 65ºC. McrBC digested or mock treated samples were PCR amplified with primers flanking 5' or 3' α*-SNAP-RAC* junctions and visualized by agarose gel electrophoresis.

RNA isolation and cDNA synthesis

Total RNAs were extracted using Trizol (Life Technologies Corp., Carlsbad CA) or the Direct-Zol RNA Miniprep Kit (Zymo Research, Irvine CA), per manufacturer's instructions. All RNA samples were DNAase treated and quantified using a spectrophotometer. cDNA synthesis was

performed using the iScript cDNA Synthesis Kit (Bio-Rad, Hercules CA) according to manufacturer's recommendations using 1.0 μg of purified total RNA.

qPCR analysis

qPCR was performed with a CFX96 real-time PCR detection system (BioRad Laboratories, Hercules CA) using SsoFast EvaGreen Supermix (Bio-Rad Laboratories, Hercules CA) as in (Cook et al., 2012). Following amplification, a standardized melting curve analysis program was performed. Overall cDNA abundances for each sample were normalized using the qPCR signal for reference gene *Glyma.18G022300*. *RAC* transcript abundances are presented relative to the mean abundance of *RAC* transcript for Williams 82 leaf samples.

RT-PCR analysis

For RT-PCR, 31 cycles of amplifications were performed prior to loading PCR product samples for separation and visualization by agarose gel electrophoresis. The number of PCR cycles terminated prior to maximal amplification of product from the most abundant template pool. A primer set sitting on a conserved copia region detected both endogenous *RAC*-like transcripts as well as the uniquely tagged *RAC* transgene. Specific detection of the tagged *RAC* transgene was with a primer pair sitting atop the engineered region. Transcript from *Glyma.18G022300* or *Skp16* served as a control for both cDNA quality and relative transcript abundance.

Vector construction

Native *α-SNAP-RAC* was PCR amplified from pTopo XL subclones using Kapa HiFi polymerase with AvrII and SbfI restriction site overhangs. Following agarose gel
(cut with XbaI/PstI) (New England Biolabs, Ipswich MA). Gel extractions were performed using the QIAquick Gel Extraction Kit (Qiagen, Hilden Germany) or the Zymoclean Large Fragment DNA Recovery Kit (Zymo Research, Irvine CA). Purified DNA fragments were ligated overnight at 4ºC with T4 DNA ligase (New England Biolabs, Ipswich MA) per manufacturer's recommendations. To remove *RAC* from within the native *α-SNAP-RAC* subclone in vector pSM101, the Polymerase Incomplete Primer Extension (PIPE) PCR method was used with Kapa HiFi polymerase (Klock and Lesley, 2009). Similarly, the synonymous tag added within the *RAC* ORF of native *α-SNAP-RAC* was created using PIPE. Unique nucleotide tag was located ~160 bp downstream of the *RAC* ATG and maintains an intact ORF. For creating the *RAC* only vector which assessed inherent *RAC* transcription, the native *RAC* ORF with both LTRs (~4.77 kb) was amplified from the initial PI 89772 *α-SNAP-RAC* subclone in pTopoXL using Kapa HiFi. Restriction site overhangs for AvrII and SbfI were incorporated into the primer sequences, and following gel recovery, the PCR amplicon was restriction digested and ligated into a PstI/XbaI cut pSM101 binary vector using T4 DNA ligase (New England Biolabs, Ipswich MA). For the native *α-SNAP-RAC* with flanking *Rhg1* sequence, an 11.1 kb native *Rhg1* sequence containing *Glyma.18G022400* and *Glyma.18G022500* (and ~1 kb downstream of each stop codon), was PCR amplified from a previously published fosmid subclone "Fos-32", with AvrII and SbfI restriction ends using Kapa HiFi polymerase (Cook et al., 2012). After restriction digestion, this amplified native fragment was ligated into the binary vector pSM101 (digested with PstI and

XbaI). This created a native *Rhg1* two gene vector of the *rhg1-b* type. Then, to make the native *rhg1-a* construct, this native *rhg1-b* pSM101 vector was used as a scaffold for step-wise cloning of two different native *rhg1-a* fragments amplified from PI 89772 genomic DNA. The first was a 4 kb fragment with an SbfI primer overhang containing *Glyma.18G022400* up until an endogenous NruI site at exon 1; the second 11 kb fragment resumed at NruI until ~ 1.0 kb downstream of the *Glyma.18G022500* (*α-SNAP$_{Rhg1}$LC*) termination codon and contained an AvrII restriction overhang.

Immunoblotting & antibodies

Affinity-purified polyclonal rabbit antibodies raised against α-SNAP$_{Rhg1}$LC and wild-type α-SNAP C-terminus were previously generated and validated in 1(Bayless et al., 2016).

Protein lysates were prepared from ~100 mg of soybean roots that were immediately flash-frozen in liquid N_2. Roots were homogenized in a PowerLyzer 24 (Qiagen) 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. Lysates were then centrifuged at 10,000g for 10 minutes and supernatant was added to SDS-PAGE loading buffer. Immunoblots were performed essentially as in (Song et al., 2015; Bayless et al., 2016). Briefly, immunoblots for α-SNAP$_{Rhg1}$LC or WT α-SNAPs 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. 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 signal detection was performed with SuperSignal West Pico or Dura chemiluminescent substrates (Thermo Scientific, Waltham WA) and developed using a ChemiDoc MP chemiluminescent imager (Bio-Rad, Hercules CA).

Immunolabeling and Electron Microscopy.

Immunolabeling was performed as in (Bayless et al., 2016). Transverse sections of ~2 mm long soybean (cv. Forrest) root areas containing syncytia were harvested by hand-sectioning at 4 dpi. Root sections were fixed in 0.1% glutaraldehyde and 4% (vol/vol) paraformaldehyde in 0.1M sodium phosphate buffer (PB) (pH 7.4) overnight after vacuum infiltration for about 1 hour. After dehydration in ethanol, samples were then embedded in LR White Resin. Ultrathin sections (~90-nm) were taken longitudinally along the embedded root pieces using an ultramicrotome (UC-6; Leica) and mounted on nickel slot grids. For the immunogold labeling procedure, grids were first incubated on drops of 50 mM glycine/PBS for 15 min, and then blocked in drops of blocking solutions for goat gold conjugates (Aurion; Wageningen, NL) for 30 min and then equilibrated in 0.1% BSA-C/PBS (incubation buffer). Grids were then incubated overnight at 4^{o}C with custom α-SNAP$_{Rhg1}$LC polyclonal antibody (diluted 1:1000 in incubation buffer), washed five times in incubation buffer, and incubated for 2 h with goat anti-rabbit antibody conjugated to 15-nm gold (Aurion) diluted 1:50 in incubation buffer. After six washes in incubation buffer and two 5-min washes in PBS, the grids were fixed for 5 min in 2.0% (vol/vol) 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. Images were collected with a MegaView III digital camera on a Philips CM120 transmission electron microscope.

Primers used in the study

RAC qRT For: GGGTTCGAAATGAATACCTG

RAC qRT Rev: CACGTTCTTCTCATGGATCCTA

RAC Delete PIPE For: CTT CAT CCA CAA TTC TAA TTT ATA TGC TAG

RAC Delete PIPE Rev: GAA TTG TGG ATG AAG TAC GAC AAT CAA C

5’ RAC Junction For: TGGCTCCAAGTATGAAGATGCC

5’ RAC Junction Rev: AACTACAGTGGCTGACCTTCT

3’ RAC Junction For: ACTGTTCATTCAGACCGCGT

3’ RAC Junction Rev: GCAATGTGCAGCATCGACATGGG

WT Junction For: GAGTTTTGAGGTGTCCGATTTCCC

WT Junction Rev: GTGAGCGCAGTCACAAACAAC

5’ Methylation For: TGGCTCCAAGTATGAAGATGCC

5’ Methylation Rev: AACTACAGTGGCTGACCTTCT

3’ Methylation For: ACTGTTCATTCAGACCGCGT

3’ Methylation Rev: GCAATGTGCAGCATCGACATGGG

Skp16 qRT For: GAG CCC AAG ACA TTG CGA GAG

Skp16 qRT Rev: CGG AAG CGG AAG AAC TGA ACC

2570 qRT For: TGA GAT GGG TGG AGC TCA AGA AC

2570 qRT Rev: AGC TTC ATC TGA TTG TGA CAG TGC

For RAC Tag Mut: GCTCTGCTCCTGAGCCCTTGAAAACGGACAGAATGCACGGAG

Rev RAC Tag Mut: GCTCAGGAGCAGAGCCATCTATGAACTCCACTTTATTCTTGGC

RAC Tag Detect For: CAGTCCTAGACTCAACCAATTACC

RAC Tag Detect Rev: CCTTGGCTATACCTGCTCTTTAAATC

For RAC initial TopoXL subclone: GAGATTACATTGGATGATACGGTCGACC

Rev RAC initial TopoXL subclone: AGATAAGATCAGACTCCAGCAACCTC

For RAC Alone subclone AvrII: cctaggGGTGTCCGATTTCCCGATTAATTGAAG

Rev RAC Alone subclone SbfI: cctgcaggCCAACATCAATTTCAAAGTTCGTCACTTTC

LC-Splice Reverse: AGTAATAACCTCATACTCCTCAAGTT

LC-Splice Full For: GAGGAGGTTGTTGCTATAACCAATGC

LC-Splice Isoform For: GAGGAGGAACTGGATCCAACATTTTC

SbfI Native *Glyma.18G022400* For: cctgcaggGAGCAGTAGGCTTCTTTGGAACTTG

AvrII Native *Glyma.18g022500* Rev:cctaggGTTCCTAAAGTGGCAAACCCTAAGAACAAAG

BglII Native For: AGATCTCCCTGAGAGTATCTTGATTTCAGATCG

BglII Native Rev: AGATCTTTTACGCATATCCGACCTTCAAC

3.6 Acknowledgements

This work was supported by USDA-NIFA-AFRI award 2014-67013-21775 and United Soybean Board award 1920-172-0122-B to A.F.B. The work was also supported by the National Science Foundation Graduate Research Fellowship under Grant No. (DGE-1256259) to A.M.B.

3.7 References

Anand SC, Gallo, K.M. (1984) Identification of additional soybean germplasm with resistance to race 3 of the soybean cyst nematode. *Plant Diseases* (USA)

Arelli PR, Sleper DA, Yue P, Wilcox JA (2000) Soybean Reaction to Races 1 and 2 of Heterodera glycines This research was partially supported by USDA-CSRS grant 95-34113-1317. Contribution of the Missouri Agric. Exp. Stn. Journal Series No. 12,751. *Crop Science* 40: 824-826

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

Bayless AM, Zapotocny RW, Grunwald DJ, Amundson KK, Diers BW, Bent AF (2018) An atypical N-ethylmaleimide sensitive factor enables the viability of nematode-resistant Rhg1 soybeans. *Proc Natl Acad Sci U S A* 115: E4512-E4521

Berg JA, Appiano M, Santillan Martinez M, Hermans FW, Vriezen WH, Visser RG, Bai Y, Schouten HJ (2015) A transposable element insertion in the susceptibility gene CsaMLO8 results in hypocotyl resistance to powdery mildew in cucumber. *BMC Plant Biol* 15: 243

Bhattacharyya MK, Gonzales RA, Kraft M, Buzzell RI (1997) A copia-like retrotransposon Tgmr closely linked to the Rps1-k allele that confers race-specific resistance of soybean to Phytophthora sojae. *Plant Mol Biol* 34: 255-264

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 Gene*t 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

Diers BW, Skorupska HT, Rao-Arelli AP, Cianzio SR (1997) Genetic Relationships among Soybean Plant Introductions with Resistance to Soybean Cyst Nematodes. *Crop Science* 37: 1966-1972

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

Dowen RH, Pelizzola M, Schmitz RJ, Lister R, Dowen JM, Nery JR, Dixon JE, Ecker JR (2012) Widespread dynamic DNA methylation in response to biotic stress. *Proc Natl Acad Sci U S A* 109: E2183-2191

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

Henk AD, Warren RF, Innes RW (1999) A new Ac-like transposon of Arabidopsis is associated with a deletion of the RPS5 disease resistance gene. *Genetics* 151: 1581-1589

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

Huang S, Tao X, Yuan S, Zhang Y, Li P, Beilinson HA, Zhang Y, Yu W, Pontarotti P, Escriva H, Le Petillon Y, Liu X, Chen S, Schatz DG, Xu A (2016) Discovery of an Active RAG Transposon Illuminates the Origins of V(D)J Recombination. *Cell* 166: 102-114

Hussey RS, Boerma HR, Raymer PL, Luzzi BM (1991) Resistance in Soybean Cultivars from Maturity Groups V-VIII to Soybean Cyst and Root-knot Nematodes. *J Nematol* 23: 576-583

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 8: 275-282

Kadam S, Vuong TD, Qiu D, Meinhardt CG, Song L, Deshmukh R, Patil G, Wan J, Valliyodan B, Scaboo AM, Shannon JG, Nguyen HT (2016) Genomic-assisted phylogenetic analysis and marker development for next generation soybean cyst nematode resistance breeding. *Plant Sc*i 242: 342-350

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

Kandoth PK, Liu S, Prenger E, Ludwig A, Lakhssassi N, Heinz R, Zhou Z, Howland A, Gunther J, Eidson S, Dhroso A, LaFayette P, Tucker D, Johnson S, Anderson J, Alaswad A, Cianzio SR, Parrott WA, Korkin D, Meksem K, Mitchum MG (2017) Systematic Mutagenesis of Serine Hydroxymethyltransferase Reveals an Essential Role in Nematode Resistance. *Plant Physiol* 175: 1370-1380

Kim MY, Zilberman D (2014) DNA methylation as a system of plant genomic immunity. *Trends Plant Sci* 19: 320-326

Klepadlo M, Meinhardt CG, Vuong TD, Patil G, Bachleda N, Ye H, Robbins RT, Li Z, Shannon JG, Chen P, Meksem K, Nguyen HT (2018) Evaluation of Soybean Germplasm for Resistance to Multiple Nematode Species: Heterodera glycines, Meloidogyne incognita, and Rotylenchulus reniformis. *Crop Science* 58: 2511-2522

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

Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* 33: 1870-1874

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, Diers BW, Hudson ME (2016) An efficient method for measuring copy number variation applied to improvement of nematode resistance in soybean. *Plant J* 88: 143-153

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

Lisch D (2013) How important are transposons for plant evolution? *Nat Rev Genet* 14: 49-61

Lisch D, Bennetzen JL (2011) Transposable element origins of epigenetic gene regulation. *Curr Opin Plant Biol* 14: 156-161

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

Makarevitch I, Waters AJ, West PT, Stitzer M, Hirsch CN, Ross-Ibarra J, Springer NM (2015) Transposable elements contribute to activation of maize genes in response to abiotic stress. *PLoS Genet* 11: e1004915

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

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

McCarville MT, Marett CC, Mullaney MP, Gebhart GD, Tylka GL (2017) Increase in Soybean Cyst Nematode Virulence and Reproduction on Resistant Soybean Varieties in Iowa From 2001 to 2015 and the Effects on Soybean Yields. *Plant Health Progress* 18: 146-155

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

McHale LK, Haun WJ, Xu WW, Bhaskar PB, Anderson JE, Hyten DL, Gerhardt DJ, Jeddeloh JA, Stupar RM (2012) Structural variants in the soybean genome localize to clusters of biotic stress-response genes. *Plant Physiol* 159: 1295-1308

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 Rhg1and Rhg4 loci. *Theoretical and Applied Genetics* 103: 710-717

Mitchum MG (2016) Soybean Resistance to the Soybean Cyst Nematode Heterodera glycines: An Update. *Phytopathology* 106: 1444-1450

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

Patil GB, Lakhssassi N, Wan J, Song L, Zhou Z, Klepadlo M, Vuong TD, Stec AO, Kahil SS, Colantonio V, Valliyodan B, Rice HJ, Piya S, Hewezi T, Stupar RM, Meksem K, Nguyen HT (2019) Whole genome re-sequencing reveals the impact of the interaction of copy number variants of the rhg-1 and Rhg4 genes on broad-based resistance to soybean cyst nematode. *Plant Biotechnol J*

Peterson-Burch BD, Voytas DF (2002) Genes of the Pseudoviridae (Ty1/copia retrotransposons). *Mol Biol Evol* 19: 1832-1845

Piya S, Bennett M, Rambani A, Hewezi T (2017) Transcriptional activity of transposable elements may contribute to gene expression changes in the syncytium formed by cyst nematode in arabidopsis roots. *Pl*ant Signal Behav 12: e1362521

Quandt HJ, Pühler A, Broer I (1993) Transgenic root nodules of Vicia hirsuta: a fast and efficient system for the study of gene expression in indeterminate-type nodules. *MPMI-Molecular Plant Microbe Interactions* 6: 699-706

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

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

Sigman MJ, Slotkin RK (2016) The First Rule of Plant Transposable Element Silencing: Location, Location, Location. *Plant Cell* 28: 304-313

Slotkin RK, Martienssen R (2007) Transposable elements and the epigenetic regulation of the genome. *Na*t 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

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

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*

Tamura K, Nei M, Kumar S (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Natl Acad Sci U S A* 101: 11030-11035

Tian Z, Zhao M, She M, Du J, Cannon SB, Liu X, Xu X, Qi X, Li MW, Lam HM, Ma J (2012) Genome-wide characterization of nonreference transposons reveals evolutionary propensities of transposons in soybean. *Plant Cell* 24: 4422-4436

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

Wawrzynski A, Ashfield T, Chen NW, Mammadov J, Nguyen A, Podicheti R, Cannon SB, Thareau V, Ameline-Torregrosa C, Cannon E, Chacko B, Couloux A, Dalwani A, Denny R, Deshpande S, Egan AN, Glover N, Howell S, Ilut D, Lai H, Del Campo SM, Metcalf M, O'Bleness M, Pfeil BE, Ratnaparkhe MB, Samain S, Sanders I, Segurens B, Sevignac M, Sherman-Broyles S, Tucker DM, Yi J, Doyle JJ, Geffroy V, Roe BA, Maroof MA, Young ND, Innes RW (2008) Replication of nonautonomous retroelements in soybean appears to be both recent and common. *Plant Physiol* 148: 1760-1771

Wickham H (2009) Ggplot2 : elegant graphics for data analysis. Springer, New York

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

Young LD (1995) Soybean Germplasm Resistant to Races 3, 5, or 14 of Soybean Cyst Nematode. *Crop Science* 35: 895-896

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. *Th*eor Appl Genet 129: 2403-2412

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

3.8 Figures

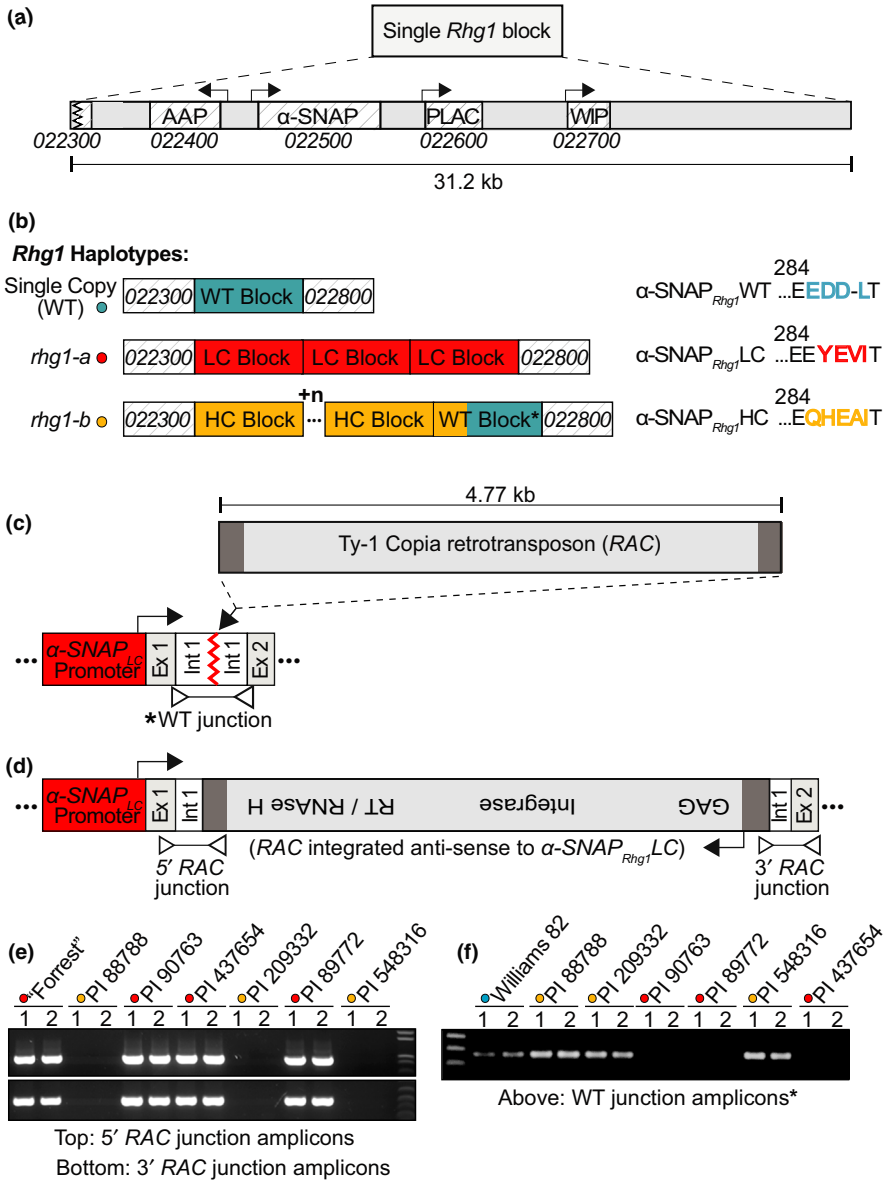


Fig 1. Multiple *rhg1-a* haplotypes harbor an intronic copia retrotransposon (*RAC*) within the *Rhg1*-encoded *α-SNAP* (*Glyma.18g022500*). (A) Diagram of a single 31.2 kb *Rhg1* block and the four *Rhg1*-encoded genes: *Glyma.18G022400* (amino acid permease, AAP), *Glyma.18G022500* (α-SNAP), *Glyma.18G022600* (PLAC-domain protein), and *Glyma.18G022700* (wound-inducible protein, WIP). *Glyma.18G022300* and *Glyma.18G022800* flank *Rhg1*, but each repeat also

includes a truncated 3' fragment of *Glyma.18G022300*. (B) Schematic of the three known *Rhg1* haplotypes: *Rhg1* wild-type (single-copy, shown blue), *rhg1-a* (low-copy, shown red) and *rhg1-b* (high-copy, shown orange). *Rhg1* α-SNAP C-terminal amino acid polymorphisms colored to match *Rhg1* block type. (C, D) Model from DNA sequencing of *Rhg1* alpha-SNAP copia (*RAC*) integration site within the PI 89772 (*rhg1-a*) encoded *α-SNAP*. The 4.77 kb *RAC* element (shown grey) is anti-sense to *α-SNAP$_{Rhg1}$LC* and increases overall repeat size to ~36 kb. *RAC* ORF is intact and encodes a 1438 residue polyprotein. *RAC* LTRs shown dark grey, *α-SNAP$_{Rhg1}$LC* promoter shown red. "Ex." and "int." are *α-SNAP$_{Rhg1}$LC* exons and introns, respectively. LTR: long terminal repeat; GAG: group-specific antigen, RT: reverse-transcriptase. (E) Agarose gel showing 5'and 3' *α-SNAP-RAC* junction products from the *rhg1-a* (low-copy, red dots) accessions: 'Forrest', PI 90763, PI 437654, PI 89772. No *α-SNAP-RAC* junctions detected from *rhg1-b* (high-copy; orange dots) accessions: PI 88788, PI 209332, or PI 548316. (F) Similar to E and using same template DNA samples, but PCR amplification of a WT (wild type) *α-SNAP* exon 1 – 2 distance, as in the Williams 82 reference genome. WT *α-SNAP* exon 1 to exon 2 distances outlined in *C* by an arrow and *.

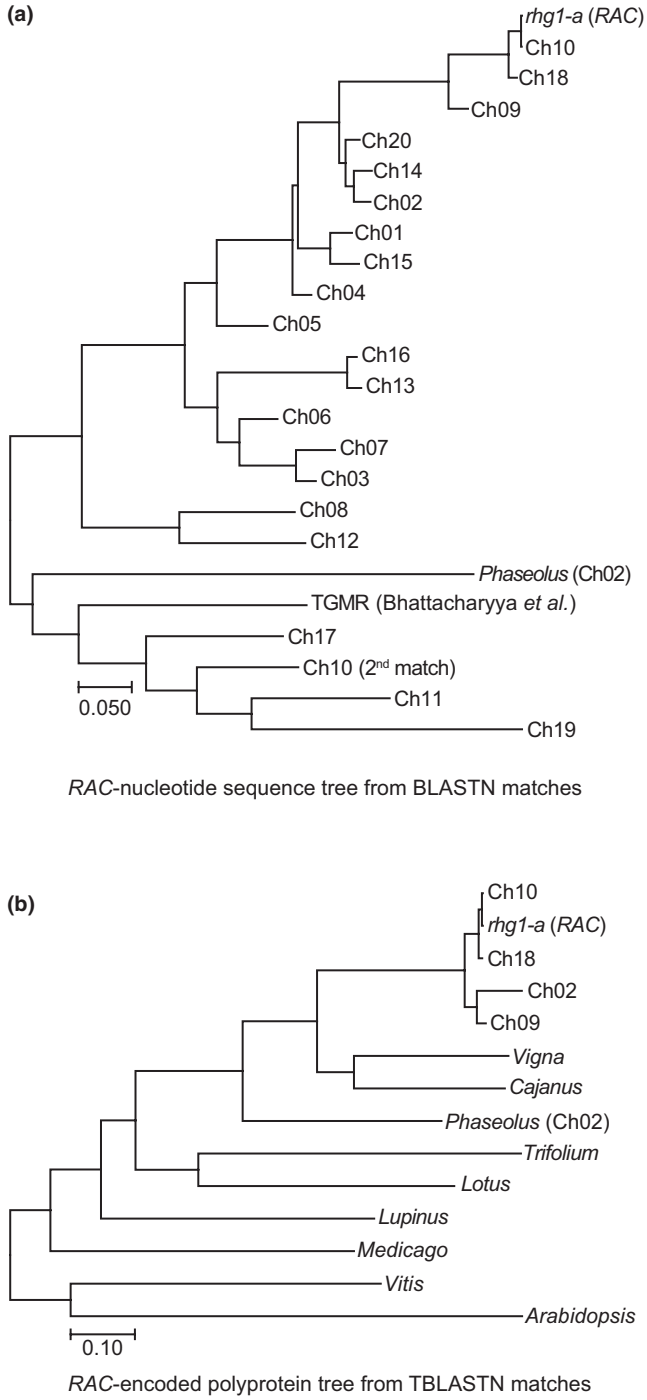


Fig 2. The *RAC*-like subfamily of copia retrotransposons is common in soybean and other legumes. (A) Maximum likelihood phylogenetic tree of *RAC*-like element nucleotide sequences from soybean. The top hit from each soybean chromosome was included, as was the known

soybean retrotransposon “TGMR” and the top *RAC*-like match from *Phaseolus* (common bean). (B) Similar to *A*; a maximum-likelihood tree, but using the *RAC*-encoded polyprotein sequences from the four most similar soybean *RAC*-like elements, and the most similar element matches from the indicated plant species.

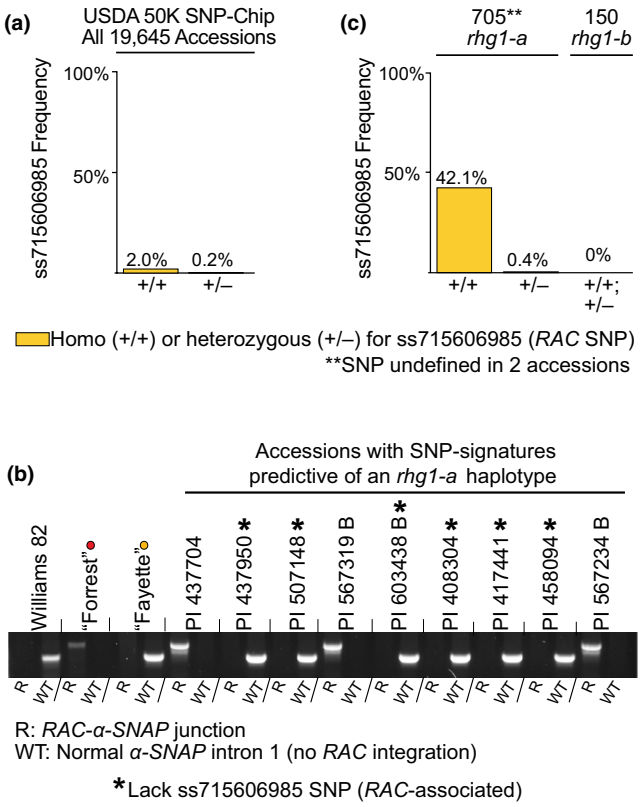


Fig 3. *RAC* is present within a subclass of *rhg1-a*-signature soybean accessions. (A) Frequency of *RAC*-associated SNP, ss715606985, among 19,645 SoySNP50K-genotyped USDA soybean accessions. (B) Agarose gel showing PCR detection of *α-SNAP-RAC* junctions or WT *α-SNAP* exon 1-2 distances among *rhg1-a*-signature accessions positive or negative for the *RAC*-SNP, ss71560698. Williams 82 ($Rhg1_{WT}$), 'Forrest' (*rhg1-a*) and 'Fayette' (*rhg1-b*) included as controls; *Rhg1* haplotypes color coded with dots as in Fig 1. An * denotes an *rhg1-a*-signature accession lacking the *RAC*-SNP. (C) Frequency of *RAC*-associated SNP among all USDA *Glycine max* accessions with consensus SNP signatures of *rhg1-a* or *rhg1-b* haplotypes.

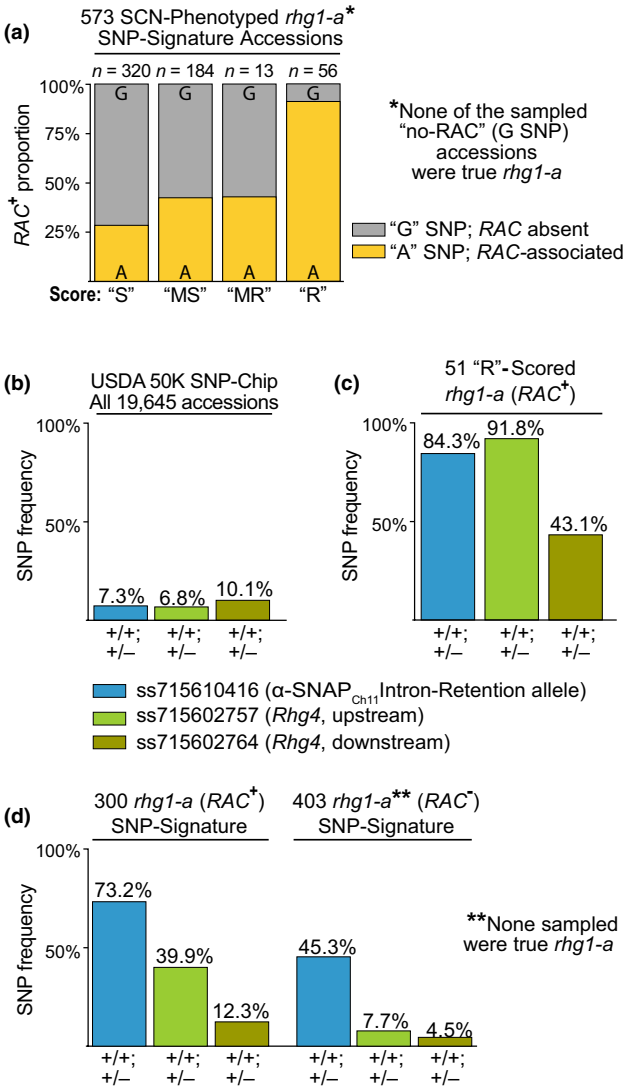


Fig 4. *RAC* presence correlates with a stronger SCN-resistance profile and the co-presence of other loci that augment *rhg1-a* resistance. (A) Proportion of *RAC*$^{+}$ (ss715606985 A SNP) vs. *RAC*$^{-}$ (G SNP) accessions among 573 SCN-phenotyped soybeans with consensus SoySNP50K SNP signatures predictive of *rhg1-a*. *Note that none of the sampled *RAC*$^{-}$ (G SNP) accessions had *rhg1-a* (none encoded α-*SNAP*$_{Rhg1}$*LC*). "S": susceptible in all trials, "MS": moderately susceptible in at least one trial, "MR": moderately resistant in at least one trial, "R": resistant in at least one trial. Fisher's Exact Test pairwise comparisons: "R-MR" (p = 2.6E-4), "R-MS" (p = 2.3E-11), "R-S" (p = 2.2E-16), "MR-MS" (p = 1.0), "MR-S" (p = 0.25), "MS-S" (p = 2.4E-3). (B)

Frequency of SNPs associated with *Rhg4* (ss715602757, ss715602764) or the Chr 11-encoded α-SNAP-Intron Retention (α-$SNAP_{Ch11}$IR) allele, ss71559743 among 19,645 USDA accessions. (C) Frequency of the *Rhg4* and *α-$SNAP_{Ch11}$IR* associated SNPs among the 51 "Resistant" scored *RAC^+ rhg1-a*-signature accessions. (D) Frequency of the *Rhg4* and *α-$SNAP_{Ch11}$-IR* associated SNPs among all *RAC^+* (300) or *RAC^-* (403) USDA *G. max* accessions with consensus SNP signatures predictive of *rhg1-a* (705 total; 2 accessions undefined for ss715606985 SNP).

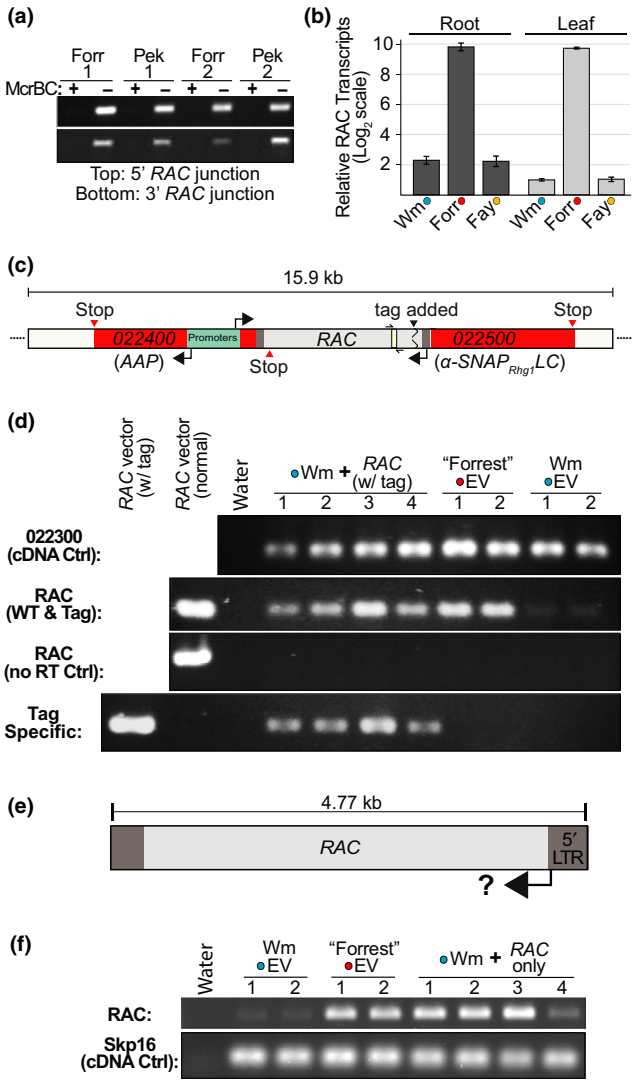


Fig 5. The *rhg1-a RAC* element is methylated but has intrinsic transcriptional activity**.** (A) Agarose gel showing PCR amplicons for *α-SNAP-RAC* regions from McrBC treated (+) or mock treated (-) genomic DNAs from 'Forrest' (Forr) or 'Peking' (Pek, PI 548402) roots. (B) qPCR analysis of mRNA transcript abundance for *RAC* and similar *RAC*-like elements, in leaf or root tissues of Williams 82 (Wm; $Rhg1_{WT}$), 'Forrest' (Forr; *rhg1-a*) or 'Fayette' (Fay; *rhg1-b*). Colored dots indicate *Rhg1* haplotype as in Fig 1. Normalized *RAC* transcript abundances are presented relative to the mean abundance of *RAC* transcript for Williams 82 leaf samples. Y-axis uses $\log_2$

scale. (C) Schematic showing unique nucleotide tag addition to an otherwise native α-*SNAP-RAC* cassette. This construct contains native flanking *Rhg1* sequence including *Glyma.18G022400* (transcribes from the bidirectional *α-SNAP* promoter) and 1.8 kb upstream, as well as 4.7 kb of downstream *RAC* flanking sequence (~1.0 kb after the *α-SNAP*$_{Rhg1}$*LC* termination codon). The *RAC* region detected and amplified via qPCR or RT-PCR is colored ivory and flanked by half-arrows. (D) Agarose gel of RT-PCR cDNAs of 'Forrest' or Wm 82 transgenic roots transformed with an empty vector (EV) or the native tagged α-*SNAP-RAC* construct. Tag primers amplify only the modified α-*SNAP-RAC* while the normal *RAC* primer set amplifies both endogenous *RAC*-like transcripts as well as the tagged α-*SNAP-RAC* transgene. *Glyma.18G022300* mRNA transcript used as a cDNA quality and loading control; no RT (reverse transcriptase) ctrl verifies absence of amplifiable genomic DNA. (E) Schematic showing the subcloned 4.77 kb *RAC* expression cassette tested in F. (F) Like D, but 'Forrest' or Wm 82 roots transformed with empty vector or the 4.77 kb *RAC* element (all flanking *Rhg1* sequence context removed).

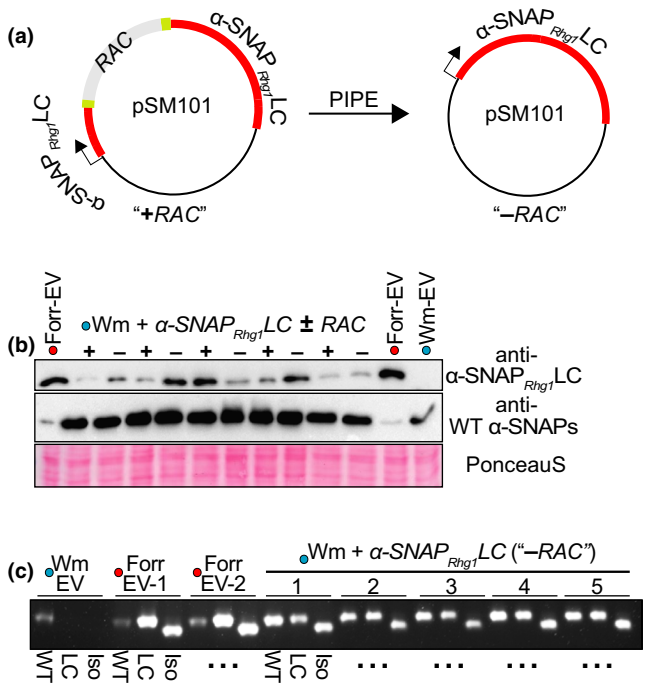


Fig 6. α-SNAP$_{Rhg1}$LC protein is expressed despite *RAC* presence. (A) Schematic showing PIPE-mediated removal of *RAC* from the native *α-SNAP-RAC* construct, pSM101. (B) Immunoblots of independent 'Forrest' or Wm82 transgenic root lysates using previously described antibodies for α-SNAP$_{Rhg1}$LC or WT α-SNAP proteins. "+" denotes *α-SNAP-RAC* transformation, "-" indicates *α-SNAP$_{Rhg1}$LC* (*RAC* removed) transformed, and EV is empty vector transformed. Ponceau S staining serves as a loading control. (C) Agarose gel showing RT-PCR amplification of mature α-SNAP$_{Rhg1}$LC transcript isoforms from roots of Wm 82 or 'Forrest' transformed with *α-SNAP-RAC* (+), or a native *α-SNAP$_{Rhg1}$LC* cassette with *RAC* removed (-), or an empty vector control. WT refers to primers specific for WT α-SNAP transcripts, LC detects full length α-SNAP$_{Rhg1}$LC transcripts, while "Iso" amplifies a previously described α-SNAP$_{Rhg1}$LC alternative transcript isoform that splices out 36 bp (Cook et al., 2014).

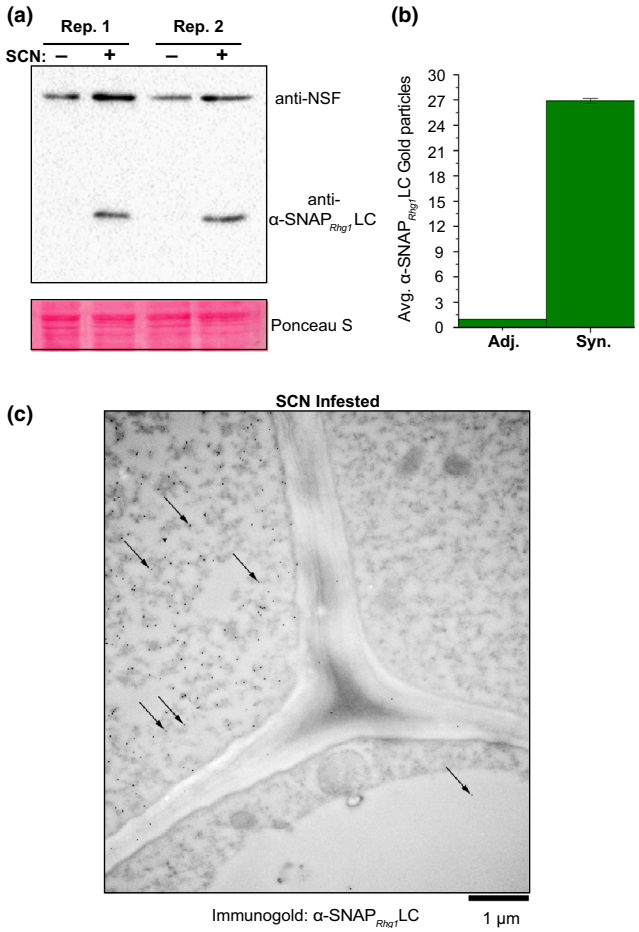


Fig 7. α-SNAP$_{Rhg1}$LC hyperaccumulates at SCN infection sites in low-copy *rhg1-a* soybean accession 'Forrest'. (A) Immunoblot of non-transgenic 'Forrest' root samples from SCN-infested root regions (SCN +) harvested 4 d after SCN infection, or similar regions from mock-inoculated controls (SCN -). Blot was probed simultaneously with anti-α-SNAP$_{Rhg1}$LC and anti-NSF polyclonal antibodies. Ponceau S staining before blotting served as a loading control. (B) Representative electron microscope image (7 dpi) showing anti-α-SNAP$_{Rhg1}$LC immunogold signal in SCN-associated syncytium cells from 'Forrest' roots. Arrows highlight only some of the 15 nm immunogold particle dots. Frequent α-SNAP$_{Rhg1}$LC signal was observed in syncytium cells (upper left, "Syn") but rare in adjacent cells (upper right and bottom, "Adj."). CW, cell wall; M, mitochondrion; Vac, vacuole. Bar = 1 µm. (C) Mean and SEM of α-SNAP$_{Rhg1}$LC gold particle

abundance in syncytia, normalized to the count from adjacent cells in the same image. Anti-α-SNAP$_{Rhg1}$LC immunogold particles were counted for one 9 μm^2 area within cells having syncytium morphology and in a region with the highest observable signal in directly adjacent cells with normal root cell morphology (large central vacuole). Data are for 23 images (11 and 12 root sections respectively, from two experiments), for root sections 7 days after inoculation.

3.9 Supplemental Information

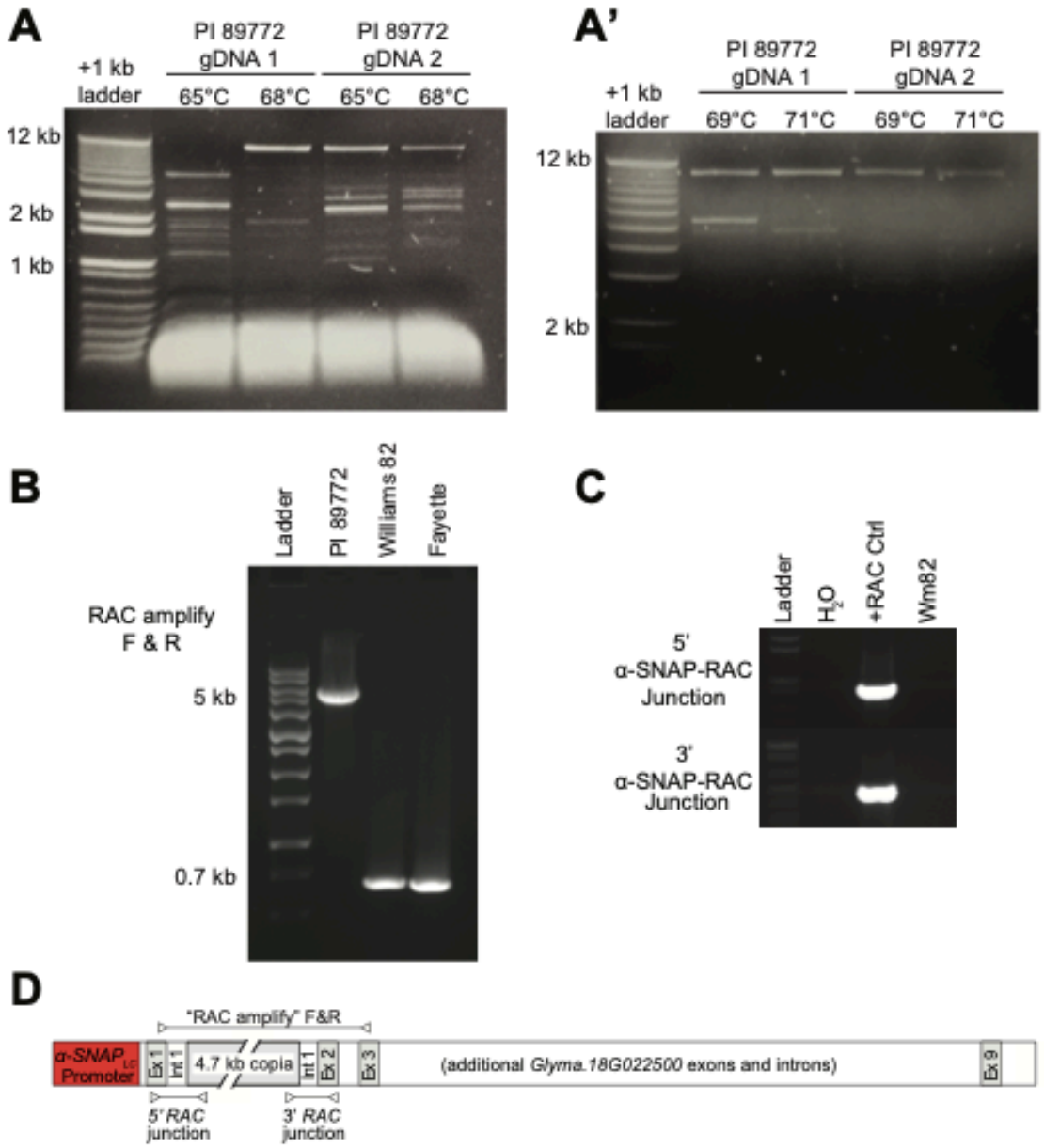


Fig. S1. Native genomic *rhg1-a* α-SNAP amplicons from PI 89772 are unexpectedly large and contain an inserted DNA element. (*A* and *A'*) Agarose gel showing PCR amplification of the native genomic *rhg1-a* α-SNAP amplicon, using a forward primer 0.85 kb upstream of transcript start and a reverse primer within region encoding 3'UTR. Similar results obtained from two independent genomic (gDNA) DNA preparations of PI 89772 and with varying primer annealing

temperatures. (*B*) Agarose gel of PCR products using primers “RAC amplify” F & R on genomic DNA from PI 89772 (*rhg1-a*), Williams 82 (*Rhg1*$_{WT}$, single copy at *Rhg1*), and Fayette (*rhg1-b*), again showing additional 4.7 kb of DNA within *rhg1-a* but not wild-type *Rhg1* or *rhg1-b*. (*C*) Agarose gel showing no detection of *RAC-*α*-SNAP* junctions within the WT *Rhg1* α-SNAP encoded by the soybean reference genome, Williams 82 (Wm82), relative to a positive control plasmid containing the subcloned PI 89772 (*rhg1-a*) *RAC-*α*-SNAP* (+RAC Ctrl); H2O - no template negative control. (*D*) Schematic of *Glyma.18G022500* within *Rhg1* locus (copia is only present in *rhg1-a* alleles), showing sites of homology for PCR primers used in (*B*) and (*C*).

α-SNAP$_{Rhg1}$*LC RAC* (*rhg1-a* associated copia) Nucleotide Sequence

Features:

α-SNAP$_{Rhg1}$LC Exon 1
α-SNAP$_{Rhg1}$LC Intron 1
RAC 3' LTR
RAC Polyprotein ORF (anti-sense to *α-SNAP$_{Rhg1}$LC* ORF)
RAC 5' LTR
α-SNAP$_{Rhg1}$LC Exon 2
Primer binding site
Putative polypurine tract (3'-GAGGGGGG-5')

ATGGCCGATCAGTTATCGAAGGGAGAGGAATTCGAGAAAAAGGCTGAGAAGAAGC
TCAGCGGTTGGGGCTTGTTTGGCTCCAAGTATGAAGATGCCGCCGATCTCTTCGAT
AAAGCCGCCAATTGCTTCAAGCTCGCCAAATCATGTTTTTCCTCTTTCTCTCTACTT
TTTTTAAATTCCATTTCGTGTCTCCTCAAAATGTTGATTTAGTGTCATAAATCATAATT
ATTATTCTCTTCTATTGTTGTTATTTTATTGTTATTACTTCAATCGACGAGTGTGTTGA
GTTTTGAGGTGTCCGATTTCCCGATTAATTGAAGTATAGTTTTAATCTGATTTTACTG
GAAAATATTTTTTTGCCTGATTTTTGTTTTTTGGAACAATTACTAGCATATAAATTAGA
ATTGTGGATGAAGTATAAGAAAACTACAGGAGCATAAGGAAGAAGAAGTGAGCTTG
AATATTTCAGAGAAGAAGATCAGCTTCAGCTACTTATTTTCGTATTAACAGAGAAGG
TTTATATACATGATGTGTGATTGTTATAACAGAAAAGCTAACTAACTCAACTAACCC
AACTACCCTTAACTGATACTGTTATACTGCTAAGAGCCCCCCTCAAGCTGGGAATG
GATATTCATCATTCCCAGCTTGTTACAGAGATGCCGAAAAACAGCAGGTGCAAGAG
CTTTGGTGAATATATCCGCTAGTTGTAAAGCTGACGAAACCGGAAGAAGCTTTAGG
AGACCTGAGTTAAGCTTTTGACGAACAATATGACAGTCTATCTCGATATGTTTAGTT
CGTTCGTGAAAAACGGGATTAGTAGCTATTTGGATGGCTGACTGATTATCACAGTA
TAAGTTCGCTGGTTGAACGAATGAGATACGAAAGTCTTGAAGCAAGAAGGTCAGCC
ACTGTAGTTCGCAAGTAGTGGAGGCAAGAGCGCGGTATTCAGCTTCGGAAGAACT
GCGTGACACAGTTGGCTGCTTCTTGGATTGCCAAGAAACCAGAGAGGAACCAAAG
TAAACTAAGTAGCCGGTAGTGGATTTCCTTGAATCTTTGCATCCAGCCCAATCCGA
GTCACTAAAGGCTCGGAGTTGTGCGGTACCTGCGGCAGTGAAGAAGATACCTGAT
CCCGGAGAACTCTTGAGGTATCGAAGAATCCGAAAGGCGGCTTGAAGATGAGCAT
TGGTGGGGGCGGCCATGTACTGGCTGAGTTGTTGAACAGCATACGTTATATCGGG
CCTGGTGTTGGTAAGGTATATTAATTTACCGATCAATCGCCGATAAGAGGAAGAAG
ACTCAGCTGAGAGAGGACTGCCCGAATCTGCCTGTAACTTCGTAGAGTAGTCTATT
GGTGTTGAATTGGGCTTGCATCCCAGCATTCCGGATGCATTTAGAATGTCTAATGT
ATACTTGCGTTGGCATAAGTGTATCCCTTTCGAGCTTCGGGCGATTTCAAGCCCAA
GGAAAAACTTTAAATCCCCAAGATCCTTGATCTTAAATTCAGAATCTAAGAGGGTGA
CAATTGTTTGTATTTCGGTCATGCTATTTCCTGTGAGAATGATGTCGTCTACATAAA
CAAGAAGGATGGTTGTGATGTTTCCAGTAAACCGCAAAAAGAGAGAGTGATCCGC

AGTTGACTGATGAAAGCCATGAGAGGTTAAGAAGCTTGACAATTTTACGAACCATT
GTCTGCTGGCTTGTTTGAGACCATAAAGAGACTTTTGAAGGCGACATACAAGCTTT
GGGTTATCAACGGAAAGTCCCGGAGGTATTTGCATATAAACCTCCTCGTCAAGTTC
TCCATGGAGGAATGCATTATTAACATCTAGCTGCCGTAAATGCCACTGGTTTAATG
CTGCAATTGCAAGAAGAAGGCGCACCGTGGTCAGCTTTGCTACCGGAGAGAAAGT
GTCAAGGTAGTCTAACCCTTCCATTTGGGTGTATCCCTTTGCAACCAGCCGCGCTT
TATGCCTTTCGATGGATCCATCTGCTCTATACTTTATTTTATAGACCCATCTGCACC
CAATAGCCGTCTTGTGAGAAGGGAGAGGTGTGAGGCGCCATGTTTGGTTCGACTG
AAGAGCTCGTAGCTCGGCTTCCATGGCCTTAATCCAGCAATCATGGCGAGAAGCA
TCGACATATGAGGTTGGCTCTGTGACGGAGGAAATATTCATGACAAAGTTCCTGTG
GGCAGGAGACAAGCGTGAGTAAGAGAGTACGGAACTAAGTGGATAACGAACAGCC
ATTGAAGTGCTTGGTGTAGAGGAAGCAAACTCTCTGTGGTAATCTCGGAGGTACGT
TGGGGTGTTTTTGGTTCTGGTGGATCGTCTAAGGTGTGAAGGAGGATGATTATGTT
CAGGTTCATTTGATGATGGTATAGAGATCATAGGTGGTGATGAAAGACTCTCTGTTT
GTGGGTCATCGTTTCTGTCGGAGAAGGATTCCGGTGAGGGAGCTGGATATTCTAA
GTGTGTATGCTGAGTTTCAGAAAGATAAGGAAAATGATCCTCATAAAATGTGATATT
TCGAGAGATGCTAACATCATTAGAGTGCAAATCATACACAAGATATCCCTTTGTATG
CATTTTAAAACCGATGAATATGCATGGATGAGCCTTAGCATCAAGCTTTTGCCGGTT
TGCCTTGAGTGTATTTATGTAACATAGACACCCGAAAACACGAAGGTTAGAAATGT
CACAAGGGTGTTTATGCAGCTTTTCATAGGGTGAAACATTATGCAAAAACGGCGTG
GGAATACAATTAATCAAGTAAGTGGCATGCGGCAAAGCGTAACACCAGAAGCTTG
GTGGTAGACTTGCCTGAAACAAAAGTGCACGTGTGACATTGAGAAGGTGCTGGTG
TTTGCGTTCTACAATTCCGTTTTGTTCTGGAGTTTCAATGCACGTGGTCTGGTGTAT
GATGCCCTTTGATGCATAGTAATGATGCATGGAGAATTCAATGCCATTATCACTTCT
GATGATCTTAACCTTGCCATCGTATTGTGTTTCAATGAATGTAATGAAGTTCATGAT
TATATGTCGGGTTTCAGCTTTGGATTTCATAAGATGAACCCATGTAAAGCGTGAGC
AATCATCAACTATAGTTAAGAAGTATTTGTGCCCATGCATGGATGGTTTAGAGCAC
GGACCCCATATGTCCATATGCAGTAAGTCAAAAATGTGAGATGCATGTGAATGGCT
AAGAGAAAAAGGTAATTTCTTTTGTTTCGCATGATGACACGTGTTGCAAACAAAATC
CTTATTATTTTTGAGAAGGGGATAGTAAGCTTTCATACATTGTATTCTTTCAGTGGAT
GGGTGGCCTAACCTAAAATGCCAAAGGTCAATAGGTATTACATTACATCGAGGGTG
AGTAATAGTGGAGTTTACGGTTTTGGTGGTCAGCTGAGCAGGTATTAAATGGTAGA
GACCGTGTTTTGCTTCAACTATACCAATCCTCATATGGCTGTTCACTTCCTGTAATA
CACACGATGTAGAGGAGAATATCAATTCACAATTAACGGAAGACACAAGTTTTGAT
ATTGAAATGATATTGAACGTAAAGGAAGGAATGTAAAGAACGTCTTGTAAAATGATG
TTTGATGAAAGCTTGACAATTCCTGAGTGGGTTGCACAAACACACTGGCCATTCGG
GAGTTTCACCGTGATAGGATCGATTTGTTTATAGGAATGAAGGTTGCGTAGGGAGT
AAGTCGCGTGGTCCGTTGCTCCTGAATCCAATATCCAGGAGACGCATGGTGTTCT
GAGAGAAAGGGACATACCTGGATTTGTCGGAGTATTGATACATGATGAGATTGAAG
CCATTTGTTTGGATTGAGAGATTGTCGTGTTTCCTGCGGATGGTTCCTGGATTAAG
GCTAGGAGTGCCTTGTACTGCTCGGGGGAGAAACGAACAGAATCATGAGCCTCGT
GGTGCTGAGCTTGATCTTCGTTAGCTTTGTTTTCAACGGCTACGAGATTGTTTACAG
TGGTTCTTCCTCCATAGGGCTTGTAACCCGGCGTGTAACCCGGTGGGTACCCGTG

TTTTCGATAACACACGTCCACAGTGTGTCCCATCTTGCCACAATGAGTGCAAGCTT
TCCTTCCACTATTCTTGTTTCTAGTATCGTGATTAGAGGGCATTCCATGTTTCTTATA
ACATGTGCTTTCTAAGTGACCAACACGTCCACAGAAGTCGCAGACAGTTTTAGCAG
CGTTTATAGAAATTTCTTTGGGTTCGAAATGAATACCTGGTCCAGCGTTTCCCAGTA
ATTGCCTTTCCTGTTGAGCCACATAGGAGAAAATCTTGGAGATAGCAGGTATAGGA
TCCATGAGAAGAACGTGGGATCGTATATTTCCGTACTGTTCATTCAGACCGCGTAG
AAACTGCATGGCCCTATCTTCGAGTTTCCGCTGTGCGATAATGGTGAATGCATTGC
AGGAACATCTTATATTACATGAGCAAATGGGATCGGGTCTAAAGTTCTCGATTTCGT
CCCAAATGACGCGTAATCGTGTGAAATACTCAGTTACTGTGAGCGTACCTTGCTTC
ATCGTCGAAGCTTCTTGTTGAAGGTCGGATATGCGTAAAAGATCTCCCTGAGAGTA
TCTTGATTTCAGATCGCGCCAAATTTCCTCGGCTTTGTCCATCCAAAGTATGCTCTG
ACGTATGGAGATGGCCACCGAATGAACTATCCACGAGACGACCATATTGTTACATC
TACGCCATGCTCCGTGCATTCTGTCCGTTTTTAGAGGTTCCGGGGCGCTGCCATCT
ATGAACTCCACTTTATTCTTGGCACTCAATGCAGTGACCATAGACCTGCTCCATGA
GTGGTAATTGGTTGAGTCTAGGACTGGGGAAACAAGAGCGATAGCTGGGTTTTCG
CTTGGATGGAGGTAGAGATAACTCTCCATGTTACTAGCAGAAGATTCGTTCATGGT
GGATAATGGAAGAATGCGCAGCAGAACTCTTCTTTTAGAAGAGCTCTGATACCATA
AGAAAACTACAGGAGCATAAGGAAGAAGAAGTGAGCTTGAATATTTCAGAGAAGAA
GATCAGCTTCAGCTACTTATTTTCGTATTAACAGAGAAGGTTTATATACATGATGTG
TGATTGTTATAACAGAAAAGCTAACTAACTCAACTAACCCAACTACCCTTAACTGAT
ACTGTTATACTGCTAAGAAAGTACGACAATCAACTCTGTGTTGTTTGTGACTACGCT
CACTTTCAATTTGACGACTAATCTCTTTATTTTGTTGAAAGTGACGAACTTTGAAATT
GATGTTGGAATAGTTCTGTTTATTGTTCTTGATTTGATCTATGTGGCATTTTAGGGG
ACAAGGCTGGAGCGACATACCTGAAGTTGGCAAGTTGTCATTTGAAG

Fig S2. Complete nucleotide sequence of the PI 89772 (*rhg1-a*, *Rhg1* low-copy) *RAC* (element and flanking exonic α-*SNAP*$_{Rhg1}$*LC* regions. Key sequence features are color coded as indicated in the above box.

S3 PI 89772 *rhg1-a RAC*-encoded polyprotein (1438 residues)

MESYLYLHPSENPAIALVSPVLDSTNYHSWSRSMVTALSAKNKVEFIDGSAPEPLKTDR
MHGAWRRCNNMVVSWIVHSVAISIRQSILWMDKAEEIWRDLKSRYSQGDLLRISDLQQ
EASTMKQGTLTVTEYFTRLRVIWDEIENFRPDPICSCNIRCSCNAFTIIAQRKLEDRAMQ
FLRGLNEQYGNIRSHVLLMDPIPAISKIFSYVAQQERQLLGNAGPGIHFEPKEISINAAKT
VCDFCGRVGHLESTCYKKHGMPSNHDTRNKNSGRKACTHCGKMGHTVDVCYRKHG
YPPGYTPGYKPYGGRTTVNNLVAVENKANEDQAQHHEAHDSVRFSPEQYKALLALIQ
EPSAGNTTISQSKQMASISSCINTPTNPGMSLSLRTPCVSWILDSGATDHATYSLRNLH
SYKQIDPITVKLPNGQCVCATHSGIVKLSSNIILQDVLYIPSFTFNIISISKLVSSVNCELIFS
STSCVLQEVNSHMRIGIVEAKHGLYHLIPAQLTTKTVNSTITHPRCNVIPIDLWHFRLGH
PSTERIQCMKAYYPLLKNNKDFVCNTCHHAKQKKLPFSLSHSHASHIFDLLHMDIWGP
CSKPSMHGHKYFLTIVDDCSRFTWVHLMKSKAETRHIIMNFITFIETQYDGKVKIIRSDN
GIEFSMHHYYASKGIIHQTTCIETPEQNGIVERKHQHLLNVTRALLFQASLPPSFWCYAL
PHATYLINCIPTPFLHNVSPYEKLHKHPCDISNLRVFGCLCYINTLKANRQKLDARAHPC
IFIGFKMHTKGYLVYDLHSNDVSISRNITFYEDHFPYLSETQHTHLEYPAPSPESFSDRN
DDPQTESLSSPPMISIPSSNEXEHNHPPSHLRRSTRTKNTPTYLRDYHREFASSTPSTS
MAVRYPLSSVLSYSRLSPAHRNFVMNISSVTEPTSYVDASRHDCWIKAMEAELRALQS
NQTWRLTPLPSHKTAIGCRWVYKIKYRADGSIERHKARLVAKGYTQMEGLDYLDTFSP
VAKLTTVRLLLAIAALNQWHLRQLDVNNAFLHGELDEEVYMQIPPGLSVDNPKLVCRLQ
KSLYGLKQASRQWFVKLSSFLTSHGFHQSTADHSLFLRFTGNITTILLVYVDDIILTGNS
MTEIQTIVTLLDSEFKIKDLGDLKFFLGLEIARSSKGIHLCQRKYTLDILNASGMLGCKPN
STPIDYSTKLQADSGSPLSAESSSSYRRLIGKLIYLTNTRPDITYAVQQLSQYMAAPTNA
HLQAAFRILRYLKSSPGSGIFFTAAGTAQLRAFSDSDWAGCKDSRKSTTGYLVYFGSS
LVSWQSKKQPTVSRSSSEAEYRALASTTCELQWLTFLLQDFRISFVQPANLYCDNQSA
IQIATNPVFHERTKHIEIDCHIVRQKLNSGLLKLLPVSSALQLADIFTKALAPAVFRHLCNK
LGMMNIHSQLEGGS*

Conserved Structural Features of Copia Polyproteins

GAG binding motif: CTHCGKMGHTVDVC

Protease motif: WILDSGATDH

Integrase (GKGY motif): GCLCYINTLKANRQKLDARAHPCIFIGFKMHTKGY

Rev. Transcriptase motif:
QLDVNNAFLHGELDEEVYMQIPPGLSVDNPKLVCRLQKSLYGLKQASRQWFVKLSSFL
TSHGFHQSTADHSLFLRFTGNITTILLVYVDDII

Fig. S3. Translation of the *RAC*-encoded polyprotein (1438 residues) from accession PI 89772.

Conserved features of the polyprotein are colored and identified as indicated in the above box.

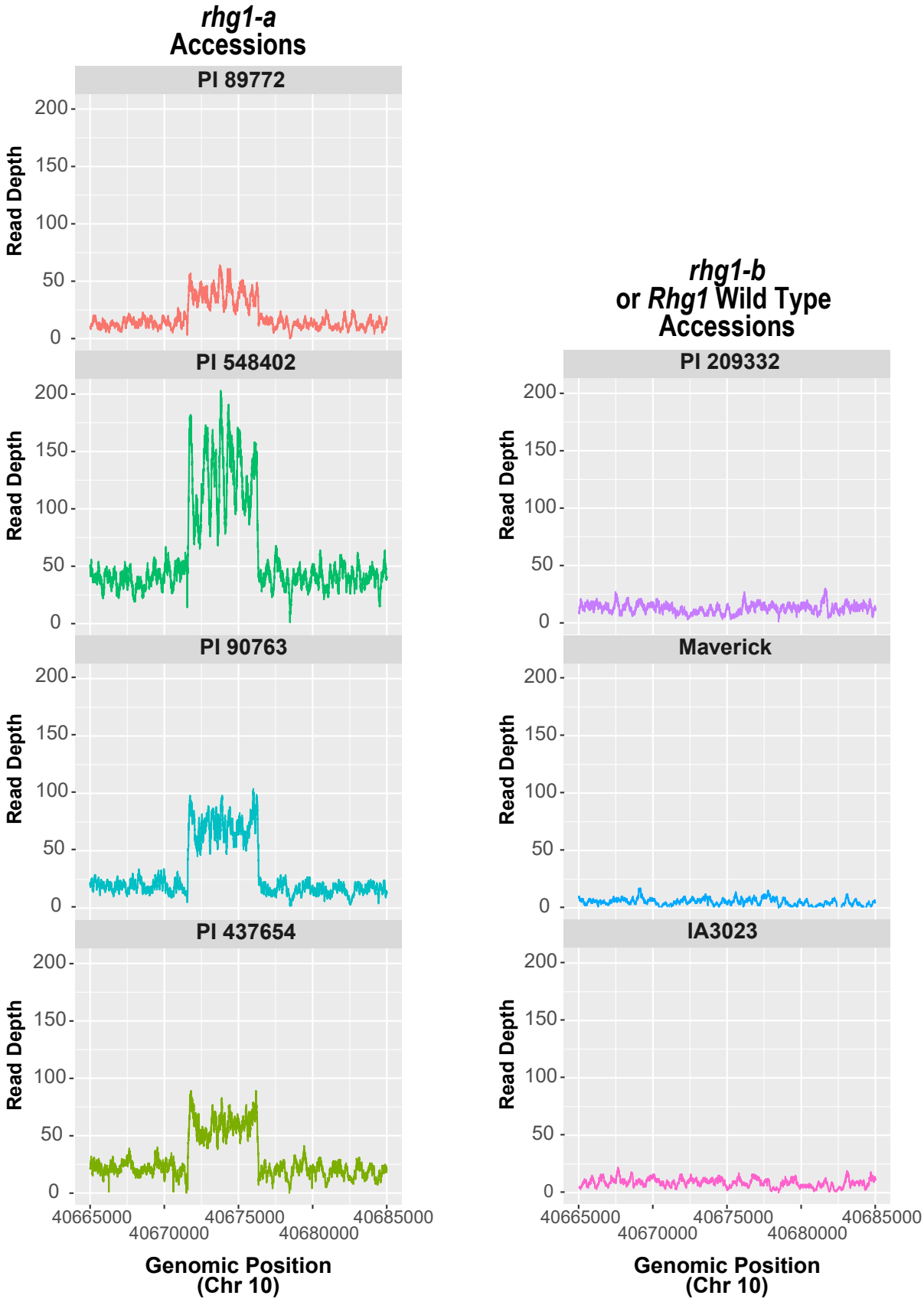

rhg1-a
Accessions
PI 89772
PI 548402
PI 90763
PI 437654
rhg1-b
or Rhg1 Wild Type
Accessions
PI 209332
Maverick
IA3023
Read Depth
0
50
100
150
200
40665000
40670000
40675000
40680000
40685000
Genomic Position
(Chr 10)

Fig. S4. Depth of short genomic reads that align to the *RAC* nucleotide sequence. The α-*SNAP-RAC* insertion is not present within the Wm 82 soybean reference genome, thus reads align to the genomic position of the Chr 10 *RAC*-like element (99.7% nucleotide identity to *RAC*). Read depth for *rhg1-a* accessions (PI 89772, PI 548402, PI 90763, PI 437654) is ~3 to 4-fold greater than for the *rhg1-b* accessions (PI 209332, Maverick) and $Rhg1_{WT}$ accession (IA3023), which do not contain Chr 18 *Rhg1* α-*SNAP-RAC* insertions.

A

```
RAC             AAGCACATGTTATAAGAAACATGGAATGCCCTCTAATCACGATACTAGAAACAAGAATAG
C10             AAGCACATGTTATAAGAAACATGGAATGCCCTCTAATCACGATACTAGAAACAAGAATAG
                ************************************************************

RAC             TGGAAGGAAAGCTTGCACTCATTGTGGCAAGATGGGACACACTGTGGACGTGTGTTATCG
C10             TGGAAGGAAAGCTTGCACTCATTGTGGCAAGATGGGGCACACTGTGGACGTGTGTTATCG
                ********************************** ***********************

RAC             AAAACACGGGTACCCACCGGGTTACACGCCGGGTTACAAGCCCTATGGAGGAAGAACCAC
C10             AAAACACGGGTACCCACCGGGTTACACGCCGGGTTACAAGCCCTATGGAGGAAGAACCAC
                ************************************************************
```

B

SoySNP 50K	*rhg1-a*	*rhg1-b*
ss715629217	C	C
ss715629233	C	C
ss715629242	C	C
ss715629248	C	C
ss715629260	G	G
ss715629264	**A**	**G**
ss715629266	T	T
ss715629273	G	G
ss715629280	G	G
ss715629286	**T**	**C**
ss715629288	**C**	**T**
ss715629291	A	A
ss715629296	C	C
ss715629298	**G**	**A**

C

Accession	*RAC* SNP?	Detected *Rhg1* α-SNAP	True *rhg1-a*?
PI 417441	No	HC α-SNAP	No
PI 507148	No	WT α-SNAP	No
PI 458094	No	HC α-SNAP	No
PI 408304	No	HC α-SNAP	No
PI 603438B	No	HC α-SNAP	No
PI 567319B	Present	LC α-SNAP	Yes
PI 567234B	Present	LC α-SNAP	Yes

Fig. S5. Sequence of the *RAC* SNP (ss715606985), nucleotide SNP signatures of consensus *rhg1-a* (*Rhg1* low-copy) and *rhg1-b* (*Rhg1* high-copy) haplotypes, and detected *Rhg1* α-SNAP transcripts from *RAC***+** or *RAC***-** accessions with consensus *rhg1-a* SNP- signatures. (*A*) Short DNA alignment of *RAC* and the Chr10 ("C10") element showing the position of the ss715606985 SNP (G to A) associated with *RAC* presence. (*B*) Consensus SNP signatures for *rhg1-a* and *rhg1-b*

haplotypes, as identified by Lee *et al*, 2015. (*C*) *Rhg1* α-SNAP alleles detected from genomic DNA subclones of *RAC***+** or *RAC***-** accessions which have a consensus *rhg1-a* SNP signature. HC, LC or WT refers to high-copy (*rhg1-b*), low-copy (*rhg1-a*) or wild-type *Rhg1* α-SNAP alleles, respectively.

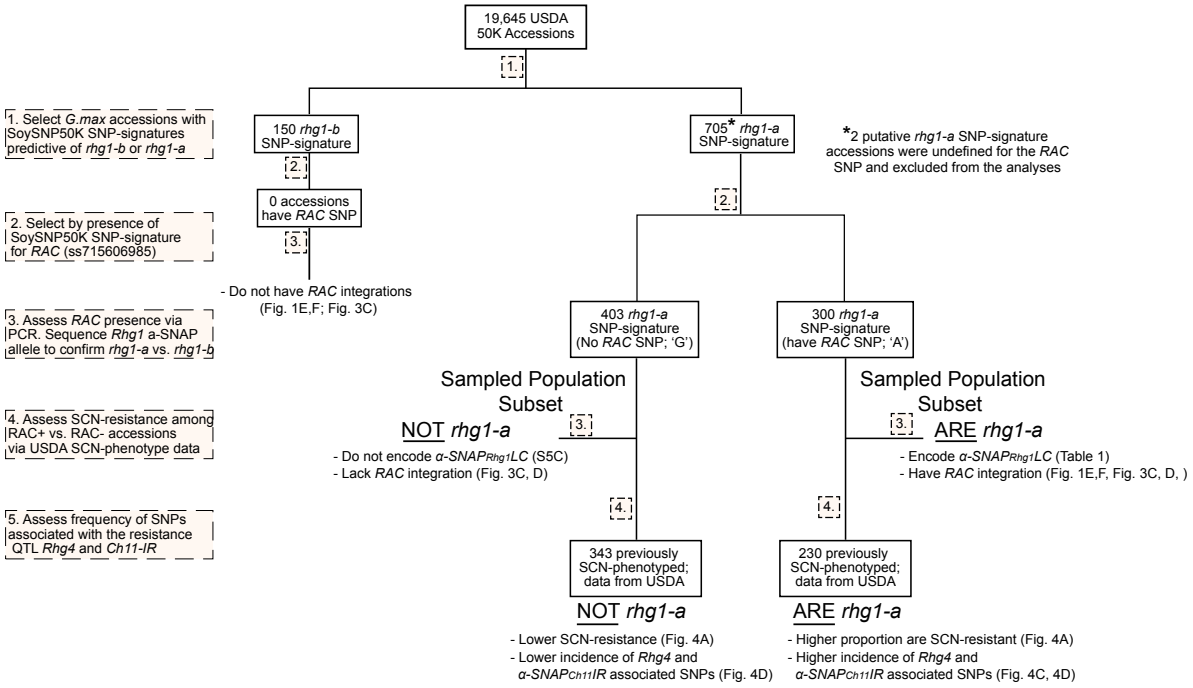


Fig. S6. Flow chart summarizing findings regarding *RAC***+** vs. *RAC***-** accessions, which otherwise have consensus SNP signatures for *rhg1-a*. The *RAC* SNP is useful to identify true *rhg1-a* accessions with strong SCN-resistance. See manuscript text for full description.

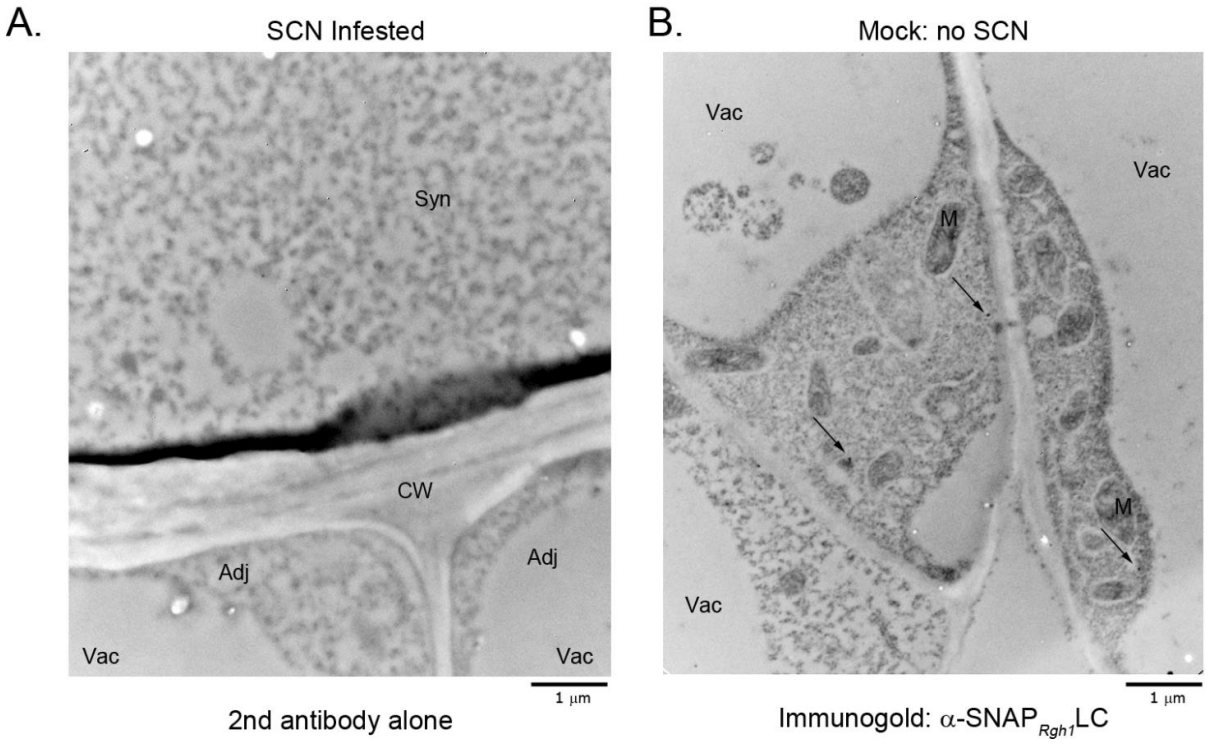


Fig. S7 α-SNAP*Rhg1*LC immunolabeling in 'Forrest' roots is highly specific. (A) Representative electron microscope image showing immunogold labeling using only secondary goat anti-rabbit antibody on SCN-infested 'Forrest' roots. Without prior incubation with the anti-α-SNAP*Rhg1*LC antibody, few gold particles were observed in syncytial vs. adjacent non-infected cells from the same root sections mounted on different grids. Syn, syncytium cells; Adj, adjacent cells. (B) Electron micrograph of a mock- inoculated 'Forrest' root after immunogold label detection using the anti–α-SNAP*Rhg1*LC primary antibody. Arrows indicate three immunogold particles across three root cell regions.

(In both A and B, CW, cell wall; M, mitochondrion; Vac, vacuole; bar = 1 µm.)

Accession	*Rhg1* Copy	SCN Resistant	*Rhg1* WT α-SNAP	*Rhg1* HC α-SNAP	*Rhg1* LC α-SNAP	WT Junction	*RAC* Junctions
Williams 82	1	–	✓	–	–	✓	–
'Forrest'	3	✓	–	–	✓	–	✓
PI 90763	3	✓	–	–	✓	–	✓
PI 437654	3	✓	–	–	✓	–	✓
PI 89772	3	✓	–	–	✓	–	✓
PI 548316	7	✓	✓	✓	–	✓	–
PI 88788	9	✓	✓	✓	–	✓	–
PI 209332	10	✓	✓	✓	–	✓	–

Table S1. Summary of features of *Rhg1* haplotypes in different soybean accessions, using data from previously published work and the current study.

Table S2

Location	BLAST Score	*RAC* Identity	Note
Ch10	8540	99.70%	Copia ORF fully intact; would encode 1438 residue polyprotein
Ch18	8076	97.60%	Integrated within intron 1 of *Glyma.18G268000* (anti-sense, copia ORF interrupted at residue 993)
Ch09	6522	90%	Copia ORF adjacent and intact; anti-sense to *Glyma.09G206300* (ABC transporter)
Ch14	4731	82%	copia ORF interrupted at residue 545
Ch02	4666	82%	copia ORF interrupted at residue 115
Ch20	4644	82%	Intronic integration within *Glyma.20G250200* (BAR domain)
Ch04	4080		
Ch01	3927		
Ch15	3910		
Ch05	3479		
Ch06	3285		
Ch03	3057		
Ch07	3036		
Ch16	2947		
Ch13	2922		
Ch08	2156		
Ch12	2040		
Ch17	1563		
Ch11	827		
Ch19	553		
P.vulgaris Ch02	378		

Table S2. *RAC*-like elements identified from NBLAST searches of *RAC* against the Williams 82 reference genome at Phytozome.org. The highest *RAC*-like match identified in the genome of the common bean (*Phaseolus vulgaris*) is also included. Only a single element (with highest identity to *RAC*) is shown for each of the n = 20 chromosomes of soybean. Several *RAC*-subfamily elements within the soybean reference genome are also inserted into host genes, and/or include elements with intact ORFs.

Chapter 4: Detection of rare nematode resistance *Rhg1* haplotypes in *Glycine soja* and a novel *Rhg1* α-SNAP

Grunwald, Derrick G.[1], Zapotocny, Ryan W.[1], Ozer, Seda[2], Diers, Brian W.[2], and Bent, Andrew F.[1]*

[1] Department of Plant Pathology, University of Wisconsin - Madison, Madison, WI USA.

[2] Department of Crop Science, University of Illinois at Urbana-Champaign, Urbana, IL USA.

*Corresponding Author: Andrew Bent, afbent@wisc.edu

Grunwald, DG., Zapotocny, RW., Ozer, S, Diers, BW. and Bent, AF. (2021). *Detection of rare nematode resistance Rhg1 haplotypes in Glycine soja and a novel Rhg1 α-SNAP*. Submitted for publication

Author Contributions

This study was conceived by D.G., R.Z. and A.B., was planned, designed and carried out by all of the authors, and was written primarily by D.G. and A.B. with contributions from all authors.

Chapter 4: Detection of infrequent nematode resistance *Rhg1* haplotypes in *Glycine soja* and a novel *Rhg1* α-SNAP

Grunwald, D.G.[1], Zapotocny, R.W.[1], Ozer, S.[2], Diers, B. W.[2], and Bent, A. F. [1]

[1] Department of Plant Pathology, University of Wisconsin - Madison, Madison, WI USA.

[2] Department of Crop Science, University of Illinois at Urbana-Champaign, Urbana, IL USA.

Corresponding Author: A. Bent, afbent@wisc.edu

4.1 Abstract

This study pursued the hypothesis that wild plant germplasm accessions carrying alleles of interest can be identified using available SNP genotypes for particular alleles of other (unlinked) genes that also contribute to the trait of interest. The soybean cyst nematode (SCN, *Heterodera glycines*) resistance locus *Rhg1* is widely utilized in farmed soybeans (*Glycine max* (L.) Merr.). The two known resistance-conferring haplotypes *rhg1-a* and *rhg1-b* typically contain 3 or 7-10 tandemly duplicated *Rhg1* segments, respectively. Each *Rhg1* repeat carries four genes including *Glyma.18G022500*, which encodes unusual isoforms of the vesicle-trafficking chaperone α-SNAP. Using SoySNP50K data for NSF_{RAN07} allele presence, we discovered a new *Rhg1* haplotype *rhg1-cs* in six accessions of wild soybean, *Glycine soja* (0.5% of the ~1100 *G. soja* accessions in the USDA collection). The α-SNAP encoded by *rhg1-cs* is unique at an important site of amino acid variation and shares with the *rhg1-a* and *rhg1-b* α-SNAP proteins the traits of cytotoxicity and altered NSF protein interaction. Copy number assays indicate 3 repeats of *rhg1-cs*. In SCN virulence assays, PI 507613 and PI 507623 have resistance to HG type 2.5.7 nematode populations, due in part to contributions from other loci. Hence the unusual multigene copy number variation *Rhg1* haplotype was present but rare in ancestral *G. soja*, and was present in accessions that offer multiple traits for SCN resistance breeding. The accessions were initially identified for study based on an unlinked SNP.

4.2 Introduction

Productive enhancements in feed and food crops are often driven by phenotypic selection during conventional plant breeding, but phenotype alone does not always reveal the donor material with the best potential for trait diversification and improvement (Tuberosa, 2012). Recent developments such as single nucleotide polymorphism (SNP) genotyping and next generation sequencing (NGS) have made it possible to prescreen germplasm for desirable traits through correlation with sequence information (Poland et al., 2012). In some cases, this approach can be extended to wild relatives of crop plants. Novel alleles of loci of agronomic importance may then be subjected to functional characterization. Alleles of interest can be moved into elite varieties via sexual crosses or transgenic approaches.

Soybean (*Glycine max*) is one of the world's most important legume crops, providing a major source of vegetable oil, protein meal for animal feed, and potential sources of renewable energy (Soystats.org; Graham and Vance, 2003; Stacey, 2010). Soybean cyst nematode (*Heterodera glycines*: SCN) is the most economically damaging pathogen of soybean, routinely causing upwards of $1 billion each year in the U.S. alone (Koenning and Wrather, 2010; Mitchum, 2016). SCN eggs can remain viable for years in cysts, which are recalcitrant to many environmental or chemical conditions, making control of established SCN populations difficult (Niblack et al., 2006). Resistant varieties and crop rotation are the major methods of SCN control. *Glycine soja* is the wild ancestor of *Glycine max*, and the more diverse gene pool in this species compared to domestic soybean offers a source of novel alleles or genes for traits of agronomic interest, including soybean cyst nematode resistance (Hyten et al., 2006; Liu et al., 2020).

The strong-effect quantitative trait locus *Rhg1* (*Resistance to Heterodera glycines 1*) is the most used source of SCN resistance, with the *rhg1-b* haplotype from PI 88788 present in the vast majority of SCN resistant soybean grown in the U.S. (Concibido et al., 2004; Donald et al., 2006; Colgrove and Niblack, 2008). Intriguingly, *Rhg1* is a tandemly repeated block of four genes, and the number of repeats varies from one in susceptible soybean varieties, or three in most *rhg1-a* haplotypes, to nine or ten copies in *rhg1-b* haplotypes (Cook et al., 2012). None of the genes in the repeated block are reminiscent of canonical defense or disease resistance genes. Both silencing and overexpression experiments have established a role in SCN resistance for three of the four genes in the *Rhg1* locus, one of which (*Glyma.18G022500*) encodes an α-SNAP (alpha-soluble N-ethylmaleimide sensitive factor (NSF) attachment protein) (Cook et al., 2012).

α-SNAP is a functionally conserved eukaryotic protein that interacts in multimeric complexes with both NSF and SNARE (soluble NSF attachment protein receptor) proteins to mediate vesicular trafficking (Jahn and Scheller, 2006). In particular, α-SNAP and NSF cooperate to promote vesicle trafficking through their disassembly for recycling of the bundled v- and t-SNARE complexes that form during vesicle fusion (Jahn and Scheller, 2006). In soybean, we recently discovered that *Rhg1*-encoded α-SNAPs are unusual in that they bind poorly to wild-type NSF (Bayless et al., 2016). In *Nicotiana benthamiana* transient assays, expression of *rhg1-a* or *rhg1-b* α-SNAPs disrupts vesicle trafficking and is cytotoxic, eventually causing cell death (Bayless et al., 2016; Bayless et al., 2018). Furthermore, these aberrant α-SNAPs accumulate ~ten-fold in syncytial cells as a response to SCN, suggesting a role of vesicle trafficking efficiency in the outcome of SCN-soybean interactions (Bayless et al., 2016; Bayless et al., 2019). Soybean

haplotypes containing three *Rhg1* repeats (low-copy haplotypes [LC]; *rhg1-a*; Peking-type or Hartwig-type) encode a distinct α-SNAP protein while those haplotypes containing nine to ten *Rhg1* repeats (high-copy haplotypes [HC]; *rhg1-b*; PI 88788-type) encode a second distinct α-SNAP protein (Cook et al., 2014). To maximize SCN resistance, varieties carrying low-copy *rhg1-a* haplotypes require presence of additional QTLs at chromosome 11 (associated with a loss-of-function intron retention allele at a chromosome 11 α-SNAP gene) and at chromosome 8 (*Rhg4*, encoding a serine hydroxymethyltransferase) (Liu et al., 2012; Lakhssassi et al., 2017; Bayless et al., 2018). These and other studies (Yu et al., 2016; Liu et al., 2017; Patil et al., 2019; Lakhssassi et al., 2020) have identified additional attributes shared by or distinct between the *rhg1-a* vs. *rhg1-b* haplotypes.

Linkage disequilibrium has been observed between the soybean chromosome 18 *Rhg1* locus and a chromosome 7 locus, with segregation distortion observed only in genotypes carrying an SCN resistance-associated *Rhg1* allele (Webb et al., 1995; Kopisch-Obuch and Diers, 2006; Vuong et al., 2015). Our group recently found that this distortion is attributable to a unique NSF allele that is encoded at *Glyma.07G195900* on chromosome 7 (Bayless et al., 2018). Termed NSF_{RAN07} (for "*Rhg1*-associated NSF on chromosome 7"), this allele is present in 11% of 19,645 soybean accessions in the USDA collection, but remarkably, is present in all soybean accessions and segregating progeny that are homozygous for the SCN resistance-associated *rhg1-a* or *rhg1-b* haplotypes. At the apparent α-SNAP/NSF binding interface, the encoded NSF_{RAN07} protein carries atypical amino acids at sites that become proximal to the unusual amino acids that distinguish α-$SNAP_{Rhg1}$LC and α-$SNAP_{Rhg1}$HC. NSF_{RAN07} protein exhibits higher

affinity than wild-type NSF for binding those resistance-associated α-SNAP proteins and enables the viability of soybeans carrying *Rhg1*-encoded SCN resistance (Bayless et al., 2018).

The gradual evolution of SCN populations toward the capacity to partially or largely overcome the SCN resistance in commercially grown soybean varieties (Niblack et al., 2008; McCarville et al., 2017) has motivated searches for new SCN resistance sources. Specific *G. soja* accessions have already provided new sources of SCN resistance that show promise (Brzostowski et al., 2017; Yu and Diers, 2017; Usovsky et al. 2020). In a recent study by Lee *et al.*(2015b), whole-genome sequencing revealed but did not further investigate a tandem duplication of *Rhg1* in *G. soja* accession Jidong 5 (W06), suggesting that resistance-conferring *Rhg1* haplotypes may have arisen prior to the divergence of *G. max* and *G. soja*. Further research on *Rhg1* duplications in *G. soja* accessions can provide insights into the evolution of SCN resistance.

In the present study, we utilized a single nucleotide polymorphism (SNP) marking the physically unlinked but genetically associated NSF_{RAN07} allele to discover and examine diversity of *Rhg1* in *G. soja* germplasm available through the USDA collection. We identified infrequent *G. soja* accessions containing tandem repeat copies of the *Rhg1* locus that represent an apparent progenitor of soybean *rhg1-a* and *rhg1-b*. We found that the α-SNAP encoded at *G. soja* multicopy *Rhg1* loci is unique, yet carried structural and functional similarities to the *rhg1-a* and *rhg1-b* α-SNAPs. Due in part to contributions from other loci, specific *G. soja* accessions carrying NSF_{RAN07} and *rhg1-cs* were found to also carry strong resistance to problematic HG type 2.5.7 SCN.

4.3 Results

A small number of *Glycine soja* accessions contain a multicopy *Rhg1* locus

As an entry point to identify *G. soja* accessions that potentially carry resistance alleles at *Rhg1*, we searched for accessions that carry genetic markers associated with the *G. max* NSF_{RAN07} allele at *Glyma.07G195900*. NSF_{RAN07} is essential for the viability of soybean plants carrying the cyst nematode resistance *Rhg1* haplotypes *rhg1-a* and *rhg1-b* (Bayless et al., 2018). The SoySNP50K Infinium BeadChip had previously been used to generate genetic data for over 19,000 domesticated soybean and over 1,000 wild soybean accessions from 84 countries, testing over 47,000 SNP loci in each accession (Song et al., 2013; Song et al., 2015). In the absence of strongly predictive *Rhg1* markers in the SoySNP50K dataset, we used SNP ss715597431 that is specific for the R_4Q polymorphism in the encoded NSF_{RAN07} protein that was recently shown to be necessary for the viability of *G. max* that carry multicopy *Rhg1* haplotypes (Bayless et al., 2018). When we queried the SoySNP50K dataset for accessions annotated as *G. soja* that contain this SNP, only 21 lines had the corresponding polymorphism, representing ~1.7% of the ~1100 *G. soja* annotated accessions (Table S1). We hypothesized that some of these accessions might also contain *Rhg1* haplotypes that confer SCN resistance.

The *Rhg1* locus was then examined in the 21 selected *G. soja* accessions. Prior work has indicated that *Rhg1*-mediated resistance in *G. max* is driven substantially by the presence of three to ten tandem duplicate copies of the >30 kb *Rhg1* locus, in so called low-copy (*rhg1-a*) and high-copy (*rhg1-b*) genotypes (Cook et al., 2012; Cook et al., 2014; Lee et al., 2015b). Lee *et al*. (2015b) noted a *G. soja* accession with three copies of the *Rhg1* locus. The vast majority of soybean accessions in the USDA collection, including the cultivar Williams 82, are SCN-

susceptible and carry the single-copy *Rhg1* haplotype (Cook et al., 2012; Lee et al., 2015b; Bayless et al., 2019). To test the hypothesis that the R_4Q-containing *G. soja* also carry multicopy *Rhg1* loci, we developed *Rhg1* repeat junction PCR primers as in Cook *et al.* (2012) but specific for *G. soja*. Primer pairs were first designed (Figure 1a) to amplify the 5' and 3' terminal portions of *Rhg1*, to confirm amplification of *Rhg1* and allow DNA sequencing of the *G. soja* products. A PCR assay was then used to test for the *Rhg1* repeat junction by pairing one outward-directed PCR primer for each terminus of the *Rhg1* repeat, which should only give a product if the 3' terminus of one *Rhg1* repeat segment is adjacent to the 5' terminus of the next *Rhg1* repeat. Screening of the 21 *G. soja* accessions revealed that six contain duplication of *Rhg1* (Figure 1b). Use of a different oligonucleotide pair, and a limited set of changes to the template preparation and PCR conditions, produced the same results regarding presence of the repeat junction PCR product in these accessions (Figure S1a).

The *Glycine soja* multicopy *Rhg1* locus shares structural characteristics with *Glycine max* multicopy *Rhg1* haplotypes

The *Rhg1* repeat junction was previously shown to be conserved between high-copy and low-copy SCN-resistant *G. max* genotypes, suggesting a shared lineage (Cook et al., 2012; Cook et al., 2014; Lee et al., 2015b). To test whether these multicopy *Rhg1 G. soja* loci also share that lineage, we cloned and sequenced the repeat junction from the six repeat-positive genotypes shown in Figure 1b. Those *G. soja* accessions were found to carry a *Rhg1* repeat junction that is identical to that of the *rhg1-a* soybean genotype Peking and the *rhg1-b* soybean accession PI 88788 (Figure 1c). As previously reported, the *Rhg1* repeat junction is not present in the single-

copy *Rhg1* of the *G. max* reference genome for cv. Williams 82 ("Wm82", Figure 1c). The above observations support the hypothesis that all known multicopy *Rhg1* types in annual *Glycine* species arose from a shared event. Due to the differences from *G. max* multicopy *Rhg1* haplotypes we describe later, we named this *G. soja* multicopy *Rhg1* haplotype "*rhg1-cs*", short for *rhg1-c* (*soja*).

Phylogenetic analyses were first conducted using genome-wide SNP data from the SoySNP50K dataset (Song et al., 2013; Song et al., 2015), to assess the overall relatedness of accessions carrying NSF_{RAN07} and *rhg1-cs*, to each other and to a more broadly representative set of *G. soja* accessions. The comparison set was a relatively random set of *G. soja* accessions chosen because they have been included in recent publications from other research groups (Figures 2a and S2a). At the whole-genome scale, *G. soja* accessions carrying an NSF_{RAN07}-encoding allele are dispersed throughout the phylogeny (Figure 2a). The rare *rhg1-cs*-containing G. soja accessions were more narrowly clustered, suggesting a shared derivation within *G. soja*. Unsurprisingly, individual rhg1-cs-containing *G. soja* also can be closely related to *G. soja* that do not carry *rhg1-cs* (Figure 2a). At a genome-wide level the rhg1-cs-containing *G. soja* group separates from G. max that carry *rhg1-a* or *rhg1-b*, suggesting that *rhg1-cs* arose independent of these haplotypes rather than from a recent hybridization with *G. max* that carry *rhg1-a* or *rhg1-b* (Figures S2a and S2b).

To study the relatedness of the *rhg1-cs* locus to the *rhg1-a* and *rhg1-b* loci, we utilized a targeted approach to extract and sequence genomic DNA from the *G. soja Rhg1* locus resulting in greater than 70% coverage of the *rhg1-cs* loci from PI 507613, PI 407287, and PI 378695A. The data were compared to previously determined *Rhg1* sequences from *G. max* and *G. soja*.

Figure 2b shows that *rhg1-cs* forms a clade with *rhg1-a* and *rhg1-b* that is distinct from the more common "wild-type" (single-copy, SCN-susceptible) haplotypes. The *Rhg1* loci from the susceptible *G. max* and *G. soja* are more similar to each other than to *rhg1-a*, *rhg1-b* or *rhg1-cs*. Although the shared origin of soybean *rhg1-a* and *rhg1-b* and *G. soja rhg1-cs* was already implied from their identical *Rhg1* repeat junction sequences (Figure 1c), the phylogeny shown in Figure 2b further demonstrates their apparent shared derivation. The data also suggest that *rhg1-cs* arose prior to the split between *rhg1-a* and *rhg1-b*.

We selected two *rhg1-cs G. soja* accessions, PI 507613 and PI 507623 (Figure 2a), to characterize in further detail. They were collected in 1983 as wild plants, about 80 kilometers from each other, in central Japan. We also continued to study *rhg1-cs*-containing PI 342434 because, although presently annotated in the USDA germplasm collection as a *G. max*, it carries leaf shape and plant architecture traits that are intermediate between *G. max* and *G. soja* and contains a genome-wide SNP pattern that clusters with *G. soja* accessions (Figure 2a). PI 342434 is also from Japan (donated to USDA in 1969) and was originally annotated as a *Glycine ussuriensis* (Regel & Maack) "tsurumame" edible wild soybean. We sought to determine the *Rhg1* copy number of these three accessions using both a genomic DNA qPCR method and a copy number variation TaqMan assay as in Lee *et al.* (2015b and 2016). Our implementation of the TaqMan assay (example shown in Figure S1b) was unsuccessful as we obtained erratic data even for controls, but those assays did indicate a *rhg1-cs* copy much lower than in *rhg1-b* cv. Fayette and similar to that of *rhg1-a* cv. Forrest. Genomic qPCR assays (Figure 1d) were more reproducible, and also included domesticated soybean cultivars of known *Rhg1* copy number 1, 3 and 10 as controls. The tested *G. soja* and *G. max* had approximately the same number of

copies of *Rhg1* as Peking (3 copies; Figure 1d). This is consistent with the hypothesis that after multicopy *Rhg1* haplotypes became established, higher soybean *Rhg1* copy number soybeans such as *rhg1-b* with nine or ten copies may have been a trait derived under positive selection during breeding (Lee et al., 2015b).

Glycine soja multicopy *Rhg1* encodes a distinct α-SNAP

Investigation of the *G. soja rhg1-cs* locus revealed that it encodes a novel *Rhg1* α-SNAP variant (Figure 3a). The contributions of *Rhg1* α-SNAP proteins to SCN resistance depend heavily on the presence of altered sets of C-terminal amino acid residues, at sites that are otherwise highly conserved across multicellular eukaryotes (Cook et al., 2014; Bayless et al., 2016). Distinct sets of C-terminal amino acids are carried by each of the *Glyma.18G022500 Rhg1* protein products: α-SNAP$_{Rhg1}$LC (from low-copy, Peking-type *rhg1-a*) and α-SNAP$_{Rhg1}$HC (from high-copy, PI 88788-type *rhg1-b*) (Cook et al., 2014; Figure 3a). The multicopy *G. soja rhg1-cs* haplotypes, and the product from *G. max* PI 342434, encoded an identical α-SNAP that diverged from the susceptible *G. max* and *G. soja* reference genomes ("wild- type" (WT) Wm82 and *G. soja* W05, and PI 483463, respectively), and from α-SNAP$_{Rhg1}$LC and α-SNAP$_{Rhg1}$HC (Figure 3a). We termed this protein α-SNAP$_{Rhg1}$Gsm, for "*G. soja* multicopy". The single- copy *Rhg1 G. soja* accessions studied because they carried the NSF$_{RAN07}$ R$_4$Q polymorphism were all found to encode an α-SNAP amino acid sequence identical to that encoded in the SCN-susceptible cv. Williams 82 soybean reference genome. Like its counterparts in established SCN-resistant soybean varieties, α-SNAP$_{Rhg1}$Gsm is polymorphic at amino acid residues predicted to form electrostatic interactions with NSF (Bayless et al., 2018). No other amino acid

polymorphisms are present between the α-SNAP of tested *Rhg1* single copy *G. soja* and α-SNAP$_{Rhg1}$Gsm. Curiously, some of the variant (non-wildtype) amino acid residues of α-SNAP$_{Rhg1}$Gsm at the C-terminus are the same as in α-SNAP$_{Rhg1}$LC while others are the same as in α-SNAP$_{Rhg1}$HC. These shared residues again suggest, as with the shared *Rhg1* repeat junction and SNP-based phylogeny, that there is a close evolutionary relationship between SCN-resistant *G. max Rhg1* haplotypes and the *G. soja rhg1-cs* haplotype.

α-SNAP$_{Rhg1}$Gsm was subsequently tested for two recently discovered resistance-associated functions. Transient expression of *Rhg1*-resistance type α-SNAPs in *Nicotiana benthamiana* is cytotoxic and induces cell death by six to seven days after infiltration with *A. tumefacien*s carrying this construct, with multiple lines of evidence suggesting that this phenotype is due to disruption of normal α-SNAP/NSF function (Bayless et al., 2016; Bayless et al., 2018). In the present experiments, introduction of a cDNA encoding α-SNAP$_{Rhg1}$Gsm driven by a constitutive promoter also caused obvious cell death in *N. benthamiana* leaves by six days after infiltration (Figure 3b). Expression of wild-type α-SNAP or an empty vector control did not cause any cell death, while expression of α-SNAP$_{Rhg1}$LC or α-SNAP$_{Rhg1}$HC caused death similar to that caused by α-SNAP$_{Rhg1}$Gsm (Figure 3b). Expression of all of the α-SNAPs was confirmed using previously described custom antibodies (Figure 3c; Bayless et al., 2016; Bayless et al., 2018).

Concurrent with cell death, overexpression of resistance-type *Rhg1* α-SNAPs has been found to cause elevated accumulation of endogenous NSF in *N. benthamiana* - an apparent cellular feedback response to disrupted NSF function (Zhao et al., 2007; Barszczewski et al., 2008; Naydenov et al., 2011; Bayless et al., 2016). Expression of α-SNAP$_{Rhg1}$Gsm caused similar

accumulation of NSF, well above that observed in control leaves expressing either the α-SNAP from cv. Williams 82 or the empty vector (Figure 3c). These cytotoxicity and NSF accumulation results suggest that, like the widely used PI 88788-type and Peking-type *Rhg1* α-SNAPs, the *G. soja* multicopy α-SNAP$_{Rhg1}$Gsm disrupts normal α-SNAP/NSF function due to its C-terminal polymorphisms.

Glycine soja Rhg1 α-SNAP interacts weakly with wild-type NSF$_{Ch07}$

In light of the above results, we investigated the interaction between α-SNAP$_{Rhg1}$Gsm and the two NSF types, wild-type NSF (NSF$_{Ch07}$) and *G. soja*-encoded NSF$_{RAN07}$. We first sought to validate the presence of *NSF$_{RAN07}$* in the six *G. soja rhg1-cs* accessions using primers specific either to the *NSF$_{RAN07}$* or *NSF$_{Ch07}$* and *NSF$_{Ch13}$* (*Glyma.13G180100*) as a positive control. As predicted by the initial SoySNP 50K screen of *G. soja* germplasm, all six of the tested accessions had *NSF$_{RAN07}$* amplicons (Figure S3). Interestingly, PI 507614B and PI 407287 gave amplicons to both *NSF$_{RAN07}$* and *NSF$_{Ch07}$* across plant samples, whereas the other accessions only gave amplicons to *NSF$_{RAN07}$*. However, sequencing of the 5' portion of *Glyma.07G195900* in PI 507614B or PI 407287 only showed sequences analogous to *NSF$_{RAN07}$*, suggesting that the *NSF$_{Ch07}$*-like locus may be adjacent to other sequences in the genome. Similarly, cloning of *Glyma.07G195900* for each of the six *rhg1-cs G. soja* accessions as well as PI 342434 '*G. max*' revealed a DNA sequence identical to that of the *NSF$_{RAN07}$* from known resistant-type soybeans. Sequencing of the 5' portion of the previously characterized PI 468396B, from Zhang *et al.* (2017), which does not have the R$_4$Q polymorphism-associated SNP, showed an *NSF$_{Ch07}$* allele. This further indicates the accuracy of the above primer set. We subsequently determined the

sequence of the *NSF$_{Ch07}$* amplicons in PI 507614B and PI 407287, by cloning the amplicon into a pGEM-T plasmid and sequencing. Intronic polymorphisms revealed that the amplicons derived from PI 507614B and PI 407287 have a sequence similar to the *NSF$_{Ch07}$* from cv. Williams 82. This suggests that there is an *NSF$_{Ch07}$* allele in these accessions, although perhaps at a chromosomal location which was not amplified from the above primer set.

Using an *E. coli* expression system, we produced recombinant α-SNAP$_{Rhg1}$Gsm protein and NSF$_{Ch07}$ and NSF$_{RAN07}$ proteins, as well as α-SNAP$_{Rhg1}$LC and α-SNAP$_{Rhg1}$HC from resistant *G. max* varieties, and α-SNAP$_{Rhg1}$WT from single-copy susceptible cv. Williams 82. We then performed *in vitro* binding assays between the different NSFs and α-SNAPs. Relative to single-copy susceptible cv. Williams 82, the resistance-associated *G. max* α-SNAPs interacted poorly with wild-type NSF$_{Ch07}$ (Figure 4a), recapitulating prior observations (Bayless et al., 2016; Bayless et al., 2018). The *G. soja* α-SNAP$_{Rhg1}$Gsm also exhibited deficient interaction with wild-type NSF$_{Ch07}$ (Figure 4a). This suggests that, as occurs with the resistance-associated soybean *Rhg1* α-SNAPs, presence of the *G. soja* α-SNAP$_{Rhg1}$Gsm is likely to disrupt the normal cellular vesicle trafficking that requires efficient α-SNAP/NSF interaction in the absence of NSF$_{RAN07}$. These results are quantified in Figure 4b. We also investigated if α-SNAP$_{Rhg1}$Gsm interacts better with NSF$_{RAN07}$ than with NSF$_{Ch07}$. Consistent with previous findings regarding α-SNAP$_{Rhg1}$LC and α-SNAP$_{Rhg1}$HC (Bayless et al., 2018), there was marked increase in the binding of α-SNAP$_{Rhg1}$Gsm to NSF$_{RAN07}$ relative to binding with wild-type NSF$_{Ch07}$ (Figures 4a and b).

Glycine soja* with multicopy *rhg1-cs* do not carry the *rhg1-a* α-SNAP copia retrotransposon (RAC), chromosome 11 α-SNAP truncation, or resistance-associated *Rhg4

The additional genetic features known to most prominently impact *Rhg1*-mediated SCN resistance are the *rhg1-a* α-SNAP copia (RAC) retrotransposon, the chromosome 11 α-SNAP intron-retention allele that encodes a non-functional truncated α-SNAP, and resistance-associated alleles of *Rhg4* that encode a serine hydroxymethyltransferase (Liu et al., 2012; Lakhssassi et al., 2017; Bayless et al., 2018; Bayless et al., 2019). We examined the six *rhg1-cs* *G. soja* accessions (and PI 342434 "*G. max*") for these genetic features. None of these seven accessions had a SNP signature of RAC integration (Table S1), or any PCR-amplification features of the 4.8 kb RAC insertion at *Glyma.18G022500* (Figure S4b). They also did not carry short sequence repeats that could be a hallmark of transposon excision at a possible excised RAC site in *rhg1-cs*.

The genomic region that in some soybean accessions encodes an intron retention allele at the *Glyma.11G234500* α-SNAP$_{Ch11}$ gene was also sequenced. All seven of the *rhg1-cs* accessions had an allele lacking the intron retention SNP (Figure S5). However, when antibodies against α-SNAP$_{WT}$ were utilized in protein immunoblots, we discovered that PI 507613, but not PI 507623, had a low apparent level of α-SNAP$_{WT}$ protein in both roots and shoots (Figure 5a and Figure S6a). This low level of α-SNAP$_{WT}$ is analogous to previous experiments with *rhg1-a* soybean lines that carry the α-SNAP intron-retention allele on chromosome 11 (Bayless et al., 2018). DNA sequencing revealed that while the encoded C-terminus of α-SNAP$_{Ch11}$ in PI 507623 is similar to cv. Williams82, the C-terminus of α-SNAP$_{Ch11}$ in PI 507613 carries a single amino acid polymorphism ($E_{284}V$; Figure 5b). The apparent low levels of α-SNAP$_{WT}$ protein in PI 507613

may be due to reduced recognition by the antibody used, or to other regulatory mechanisms, but apparently are not due to a *Glyma.11G234500* intron retention allele.

We also performed immunoblots with leaf and root extracts of PI 507613 and PI 507623 using other antibodies, to detect resistance-associated α-SNAP$_{Rhg1}$LC and α-SNAP$_{Rhg1}$HC. For tissue from developing trifoliates, immunoblots revealed a low basal expression of α-SNAP$_{Rhg1}$Gsm in *G. soja* PI 507613 and PI 507623 relative to the cv. Forrest (*rhg1-a*) control (Figure S6a). However, in *rhg1-cs* roots we observed more α-SNAP$_{Rhg1}$Gsm protein, at levels analogous to cv. Forrest soybean roots (Figure 5a). Control blots of purified α-SNAP$_{Rhg1}$Gsm protein showed antibody recognition at sensitivities similar to that for α-SNAP$_{Rhg1}$HC or LC, against which the antibodies were raised (Figure S6b).

Rhg4 encodes a variant serine hydroxymethyltransferase in the allele (not genetically linked to *Rhg1*) that plant breeders pair with *rhg1-a* to achieve useful SCN resistance in domesticated soybean (Liu et al., 2012; Yu et al., 2016; Patil et al., 2019). Sequencing of *Rhg4* cDNA clones from the six *rhg1-cs G. soja* accessions (as well as soybean *Rhg4* transcripts from SCN-susceptible cv. Essex and SCN-resistant cv. Peking as controls) revealed absence of the R_{130}P and Y_{358}N polymorphisms associated with Peking-type resistance, suggesting that the six *rhg1-cs G. soja* accessions carry a susceptible-type (wild-type) *Rhg4*.

The *Glycine soja* PI 507613 and PI 507623 are resistant to virulent nematode populations

In light of the above observations about *rhg1-cs*, including observations about the absence of a resistance-associated *Rhg4* and chromosome 11 *α-SNAP* alleles, we sought to determine the SCN resistance of select *rhg1-cs G. soja* accessions. The *rhg1-cs G. soja* lines PI

507613, PI 507623 and the "*G. soja*-like" PI 342434 were tested with HG type 0 and HG type 2.5.7 nematode populations. We compared the female indices (relative number of cysts formed) to those for *G. max* HG type indicator lines and the *G. soja* PI 468916 which is the source of the SCN resistance QTL *cqSCN-006* and *cqSCN-007* (Wang et al., 2001) (Table 1).

As might be predicted due to the absence of resistance-associated *Rhg4* or chromosome 11 α*-SNAP* alleles, PI 342434 was only able to confer moderate resistance to an HG type 0 SCN population. Intriguingly, PI 342434 was not able to confer substantial resistance to an HG type 2.5.7 SCN population. Even more intriguing, PI 507613 and PI 507623 were both able to confer strong resistance to HG type 0 and HG type 2.5.7 nematode populations, with levels comparable to PI 468916 (Table 1). The above findings, taken as a whole, do not demonstrate causation but do suggest that the novel *rhg1-cs* haplotype is as functional as *rhg1-a* for SCN resistance against HG type 0 SCN. The resistance of PI 507613 and PI 507623 to the HG type 2.5.7 SCN population (that PI 342434 was susceptible to) indicates that in PI 507613 and PI 507623, loci other than *rhg1-cs* also contribute to SCN resistance against economically important HG type 2.5.7 SCN populations.

To test the effect of the novel *rhg1-cs* allele on resistance relative to the *rhg1-b* allele, a F_2 population from a cross of *G. max* LD10-10198 (*rhg1-b* resistance) X *G. soja* PI 507623 (*rhg1-cs*) was tested for SCN resistance. Each plant in the population was evaluated for resistance in a greenhouse against an HG type 2.5.7 SCN population, and tested with the marker Satt309 which maps with 1 cM of *Rhg1* (Kim et al., 2010). A month after the plants were inoculated the susceptible check Lee 74 averaged 219 SCN females while the 96 F_2 plants averaged 120 females. Within the F_2 population, no significant association was observed between the Satt309

allele present and level of SCN resistance (single factor ANOVA Pr(>F) = 0.34). Hence no distinction was detected between *rhg1-b* and the *rhg1-cs* haplotype derived from PI 507623 in their contribution to resistance against the tested HG type 2.5.7 SCN population. This again suggests that there are loci other than *Rhg1* in PI 507623 that also contribute to the observed elevated resistance to the virulent HG type 2.5.7 SCN population (Table 1).

4.4 Discussion

Crop wild relatives are an appealing source for novel traits of agronomic interest, and this is evident in domesticated soybean. Recent resequencing of *G. soja* along with landrace and elite *G. max* varieties revealed that many rare alleles were lost from *G. soja* during its domestication to form *G. max* (Hyten et al., 2006; Liu et al., 2020). Many of these lost traits are of agronomic importance. QTLs associated with improved yield, salt tolerance, alkaline salt tolerance, soybean aphid resistance, foxglove aphid resistance, and SCN resistance have all been found to be present in *G. soja* but absent in *G. max* (Wang et al., 2001; Concibido et al., 2003; Kabelka et al., 2005; Kabelka et al., 2006; Lee et al., 2009; Hesler et al., 2013; Kim et al., 2013; Lee et al., 2015a; Yu and Diers, 2017). With respect to SCN, *G. soja* PI 468916 has served as a donor for the QTLs *cqSCN-006* and *cqSCN-007* that each contribute modest resistance to a highly virulent SCN population, but provide substantial and synergistic resistance even to HG type 1.2.3.5.6.7 SCN when combined in *G. max* with *rhg1-b* and a chromosome 10 QTL from PI 567516C (Brzostowski et al., 2017). Here, we used a SNP marker associated with NSF_{RAN07} to select *G. soja* accessions to screen for multicopy *Rhg1* haplotypes, leading to discovery of the *rhg1-cs* haplotype that encodes a unique α-SNAP with SCN resistance-associated functional

features. The hypothesis was that plant germplasm accessions that are strong candidates to carry alleles of interest can be identified using available SNP genotypes for particular alleles of other (unlinked) genes that also contribute to the trait of interest. Our study highlights the power of deploying new findings regarding the genetic architecture of traits of interest (such as the requirement for NSF_{RAN07} in *Rhg1*-mediated SCN resistance), along with available data such as the SoySNP50K data for ~20,000 *Glycine* accessions, to prescreen germplasm and identify a manageable number of accessions for functional analysis. Recent NGS work has further identified structural and nucleotide variants within domesticated soybean (Torkamaneh et al., 2019). The same study also identified loss of function alleles within soybean germplasm, potentially functioning as a sequence-catalogued mutant library analogous to those available in *Arabidopsis*. If genetic features of potential interest are known, there are increasingly powerful resources available for *in silico* germplasm prescreens.

A potential limitation of using the NSF_{RAN07} SNP for *in silico* germplasm screens is that it may not fully capture the diversity of *Rhg1*-containing *G. soja*. There may be lines that do not share with soybean (Bayless et al., 2018) the requirement for NSF_{RAN07} for *Rhg1*-mediated SCN resistance. However, we and others have identified SoySNP 50K (or KASP) SNPs associated with other resistance-associated genetic elements that might be used in such screens. In addition to NSF_{RAN07}, these includes the chromosome 11 intron retention α-SNAP knockout, the RAC (*rhg1-a*-alpha SNAP copia) retrotransposon in the first intron of *rhg1-a* α-SNAP genes, and SNPs specific for SCN resistance-associated alleles of *Rhg4* (Matsye et al., 2012; Shi et al., 2015; Bayless et al., 2018; Bayless et al., 2019; Tran et al., 2019). These might be used to select *G. soja* accessions that can subsequently be tested for the presence of useful alleles at other loci that

contribute to SCN resistance, whose presence is not adequately tested by the SoySNP50K dataset. More direct tests using markers not on the SoySNP50K chip could also be carried out *de novo* on large numbers of germplasm accessions (albeit at appreciable new expense), for example to detect *Rhg1* tandem repeat junctions, novel C-terminal sequences in α-SNAP proteins, or new *Rhg4* alleles.

The present discovery of PI 507613 as an accession of interest for SCN resistance provides an extended example of the above hypothesis, that useful alleles at multiple loci controlling a trait are likely to co-occur in individual accessions. Phenotypic screening of USDA accessions for SCN resistance apparently missed PI 507613, which was instead flagged for further study by our NSF_{RAN07} genotypic screen. PI 507613 would have been passed over in a screen for RAC but would be positive in a screen for the *Rhg1* tandem repeat junction. PI 507613 would also be negative for *Rhg4* or the chromosome 11 intron-retention α-SNAP, but we did discover another α-SNAP with unusual C-terminal amino acids in this accession. The example gains more interest when considering that, in addition to finding the novel *rhg1-cs*, we found that PI507613 carries the valuable trait of SCN resistance to HG type 2.5.7 SCN (that apparently is not dependent on traditional *rhg1-a*/*Rhg4*/chromosome 11 α-SNAP genotypes). This PI 507613 SCN resistance is dependent on more loci than just *rhg1-cs*. We postulate that it was much more likely to identify such an accession once we focused on *G. soja* accessions that carry one or more SCN resistance alleles of interest. A productive target for future studies will be to map and identify the full complement of other loci that contribute to resistance to HG type 2.5.7 SCN in PI 507613, PI 507623, and possibly in the other *rhg1-cs* PIs identified in this study.

Unlike domesticated soybean, *G. soja* accessions are not known to contain the resistance-associated allele of *Rhg4*, which is necessary for the efficacy of SCN resistance conferred by *rhg1-a* in *G. max* (Meksem et al., 2001; Brucker et al., 2005; Wu et al., 2016; Yu et al., 2016). Our findings are consistent in these observations. The accessions we examined did not have a resistance-type *Rhg4* or other attributes of Peking-type SCN resistance: presence of RAC or the non-functional chromosome 11 α-SNAP allele (although in PI 507613 there does seem to be a novel α-$SNAP_{Ch11}$). Previous GWAS studies with *G. soja* have either not detected significant contributions from the *Rhg1* locus when testing with a Race 1 (HG type 2.5.7) SCN population (Zhang et al., 2017) or detected only a minor contribution (R^2 = 13.6%) from a locus 5-12 Mb (>8 genes) away from *Rhg1* when testing with a separate HG type 2.5.7 SCN population (Zhang et al. 2016**)**. Those GWAS results may have been impacted by a very low rate of occurrence of multi-copy *Rhg1* loci in *G. soja*. However, they are also consistent with poor expression of the SCN resistance phenotype that might be predicted if a haplotype such as *rhg1-cs* is not coupled with complementary *Rhg4*/*RAC*/chr11 α-SNAP genotypes. This makes it all the more interesting to understand the other genetic features that contribute to any observed SCN resistance in *G. soja* (possibly including loci identified by Kabelka et al, 2005; Winter et al., 2007; Arelli et al, 2010, Vuong et al. 2010, Zhang et al., 2016; Zhang et al., 2017; and other studies). Presumably, novel cellular mechanisms that could enhance the durability of SCN resistance are encoded at those loci.

The evolutionary history of the economically important *rhg1-b* and *rhg1-a* soybean loci is a point of interest, and not only because of the unique structure of these loci (copy number variation of up to ten copies of a four gene chromosomal block, in which three of the four

tightly linked genes contribute to SCN resistance). Investigations of *Rhg1* evolution may also reveal new *Rhg1* haplotypes or suggest approaches to additional functional engineering of *Rhg1* components. The previous cloning and characterization of both *rhg1-a* and *rhg1-b* demonstrated that the bridge junction (the junction between two *Rhg1* blocks) is conserved (Cook et al., 2012). This suggested that *Rhg1* duplication preceded the split between *rhg1-a* and *rhg1-b*. Further studies have revealed an array of copy numbers for *rhg1-a* and *rhg1-b* haplotypes in various soybean accessions (Cook et al., 2012; Cook et al., 2014; Lee et al., 2015b; Lee et al., 2016; Yu et al., 2016; Patil et al., 2019). Whole genome resequencing had suggested the existence of a *rhg1-a*-like *Rhg1* within *G. soja*, providing evidence for an early origin of *Rhg1* outside of *G. max* (Lee et al., 2015b). The data we present provides more complete evidence of an early origin of multicopy *Rhg1*, and revealed a novel haplotype and novel α-SNAP$_{Rhg1}$. Further, this study identified *rhg1-cs* as a candidate progenitor of *rhg1-a* and *rhg1-b*, and α-SNAP$_{Rhg1}$Gsm as a candidate progenitor of α-SNAP$_{Rhg1}$LC and α-SNAP$_{Rhg1}$HC. The multi-year undertaking has been initiated to generate transgenic soybean lines in which isogenic presence/absence of the allele encoding α-SNAP$_{Rhg1}$Gsm can be associated with resistance to different HG types of SCN.

Vesicle trafficking is essential to eukaryotic cellular homeostasis. The aberrant α-SNAP encoded by resistance-associated *Rhg1* haplotypes has been implicated in conferring resistance to SCN (Cook et al., 2012; Cook et al., 2014; Bayless et al., 2016; Liu et al. 2017; Lakhssassi et al. 2017; Lakhssassi et al., 2020). Interestingly, the only SNPs causing non-synonymous codon differences in the *Rhg1* proteins from resistant and susceptible cultivars occur within the gene body of the α-SNAP, particularly at the C-terminus—the region that stimulates NSF activity for

SNARE disassembly (Barnard et al., 1997; Marz et al., 2003; Cook et al., 2014; Zhao et al., 2015; Bayless et al., 2016). Here, we report an additional resistance-associated α-SNAP, with a C-terminus distinct from that of either known resistant types or the standard wild-type α-SNAP. Variation in nematode functions that interface with host plant α-SNAPs might be associated with SCN virulence. Recent work elucidating effectors of SCN suggests that SNARE-like proteins—proteins that normally interact with α-SNAP and NSF to form the 20S complex—constitute a large and variable effector family in SCN (Gardner et al., 2018). Further, there is some molecular evidence for direct physical interaction between SNARE-like protein effectors and α-SNAP with virulence outcomes (Bekal et al., 2015). Moreover, recent evidence suggests a potential physical interaction between the *Rhg4*-encoded serine hydroxymethyltransferase and the α-SNAP$_{Rhg1}$, as well as α-SNAP$_{Rhg1}$ and syntaxin proteins, both with implications for resistance (Dong et al., 2020; Lakhssassi et al., 2020). In light of this, varying the soybean α-SNAP repertoire (for example by use of *rhg1-cs* and α-SNAP$_{Rhg1}$Gsm) may hinder the capacity of nematodes to overcome *Rhg1*-mediated resistance.

The present study also discovered a novel α-SNAP on chromosome 11 in PI 507613. From previous work it is apparent that SCN resistance involves not just presence of a α-SNAP$_{Rhg1}$, but also modification of the relative amount of wild-type α-SNAP protein, in syncytia or in whole plants (Bayless et al., 2016; Bayless et al., 2018; Bayless et al., 2019). It is possible that the novel α-SNAP$_{Ch11}$ of PI 507613 is a bona fide resistance-associated QTL. Indeed, under a model where HG type 0 or HG type 2.5.7 nematode populations produce effectors that interact with α-SNAP$_{WT}$ towards facilitating infection or reproduction, modifying the α-SNAP$_{WT}$

could change virulence outcomes due to reduced interaction. Future work should be directed towards understanding whether the novel α-SNAP$_{Ch11}$ co-segregates with SCN resistance.

It may be possible to engineer soybeans with novel α-SNAPs by genetic modification and gene-editing methods, or by classical breeding to bring in different α-SNAPs such as the *G. soja rhg1-cs* product. Through work such as that of Marz et al. (2003), a library of mammalian α-SNAP mutations and known consequences is already available, and this type of structure/function knowledge (see also Zhao et al. 2015) is ready to be translated to soybean α-SNAP. Vesicle trafficking has been implicated in many plant-pathogen interactions beyond SCN (Hoefle and Hückelhoven, 2008; Inada and Ueda, 2014). Due to the seeming centrality of vesicle trafficking in plant defense, α-SNAP protein sequences and expression patterns may also represent appealing targets to modify in order to confer resistance to other biotrophic pathogens.

4.5 Materials and Methods

Plant materials

Seeds of 22 *Glycine soja* or *Glycine max* (the first 22 entries in Table S1) were obtained from the USDA Soybean Germplasm Collection in Urbana, Illinois (https://www.ars-grin.gov/). LD10-10198 is an elite *G. max* experimental line developed at the University of Illinois that carries the *rhg1-b* haplotype for SCN resistance from PI 88788

DNA extractions and oligonucleotide primers

Soybean genomic DNA was extracted from young cotyledons from respective *G. soja* or *G. max* accessions using standard CTAB extractions as in Keim *et al*. (1988) with modifications from Cook *et al.* (2012). The oligonucleotide primers used for PCR and other assays are listed in Table S2.

Copy number variation assays

Copy number variation assays were performed essentially as in Lee et al., 2015b. In brief, 80 ng of extracted genomic DNA was subjected to qPCR using primers directed against the bridge junction as in Figure 1 or *Glyma.18G022800* using SolisBiodyne 5X Firepol (SolisBiodyne; Cat. No. 08-36-00001). Copy number was estimated using the ΔCt method where ΔCt is the difference between *Glyma.18G022800* and bridge amplifications. For TaqMan assays, experiments were performed essentially as in Lee *et al*., (2016).

Targeted DNA extraction

DNA was enriched for the *Rhg1* locus using 837 specific patch oligos as in Varley *et al.,* 2010, with significant modifications. DNA was first digested using three combinations of restriction enzymes: (1) DraI/TaqαI/NlaIII, (2) Mn1I/MboII, and (3) ApoI/HpyCH4V (100 ng DNA per reaction). A patch driven ligation was performed with 2 nM of each complementary oligo (Table S2), 40 nM Nested Forward, 20 nM 'U2', 20 nM 'U2-Phospho', 5 U of Ampligase (Epicentre Cat. No. A3202K) and 1X Ampligase Buffer in 25 μl. This reaction was incubated for 10 min at 95°C, followed by 30 sec at 95°C and 4 min at 54°C for 80 cycles. Undesired products and excess

genomic DNA was then digested with the direct addition of 5 U Exonuclease I (USB) and 100 U Exonuclease III (Epicentre Cat. No. EX4425K). This was incubated at 37°C for 1 hour followed by inactivation at 95°C. Products were purified using 42 μl (1.6X) of Agencourt AMPure XP beads (Beckman Coulter Cat. No. A63880) and eluted in 20 μl ddH20. A nested PCR was then performed using KAPA HiFi HotStart Uracil+ Readymix (2X) (KapaBiosystems Cat. No. KK2801), 100 μM of the oligos 'Right-no homology' and 'Left-no homology', 2.5 mM $MgCl_2$ and 20 μl of the eluted product from previous step. This reaction was incubated according to manufacturer instructions with an annealing temperature of 66°C and 30 second extension for 31 cycles. Product was again purified using 1.6X sample volume of Agencourt AMPure XP beads before submission for sequencing at the University of Wisconsin Biotechnology Center (UWBC). Samples were individually barcoded before being sequenced using Illumina MiSeq (Illumina) 2X250 with 50% PhiX.

Sequence variant detection

Flow cell data were demultiplexed and Illumina adapter sequences trimmed by the UWBC. Universal sequences were trimmed using Cutadapt (version 2.8; Martin, 2011). Previously published whole genome sequencing read data from soybean HG type indicator lines as well as the *G. soja* genotypes W05 were also added to the analysis moving forward. (Cook et al., 2014, Bioproject PRJNA243933 (Hg Type Test Lines); Xie et al., 2019, Bioproject PRJNA486704(W05)). Full genome alignment was performed as in Cook *et al.*, 2014 using the Wm82.a2.v1 reference genome (Schmutz et al., 2010) utilizing the programs BWA (version 0.7.17, Li and Durbin 2010) and Picard (version 2.19). The HaplotypeCaller function with the GATK software package (Poplin

et al., 2017) was used to detect variants in the genomic region Gm18:1632666-1663881. Due to the nature of our targeted DNA extraction, we lacked complete coverage of the locus. Because of this we identified regions within *Rhg1* where all the sequenced *G. soja* samples had depth in excess of three. Vcftools (version 1.13, Danecek et al., 2011) was used to pull out variants within these regions and filter for a depth of three and variant quality value of 50. A fasta sequence file was obtained for each sample using the FastaAlternateReferenceMaker function within GATK. These fasta files were then imported to SplitsTree for network analysis (Huson and Bryant 2006).

Network Analysis and Phylogenetic Trees

To determine the relationship between *G. soja* and *G. max* or among *G. soja*, we aligned all the SNPs available in the SoySNP50K database (Song et al., 2013; Song et al., 2015). The program SplitsTree (v 4.16.1) was used to perform the alignment and construct the network (Huson and Bryant, 2006). Uninformative sites were excluded in the network construction. Rooted phylogenetic trees were built using MEGAX (Kumar et al., 2018; Stretcher et al., 2020).

RNA isolation and cDNA synthesis

Total RNA was extracted from expanding soybean trifoliates using either the DirectZol RNA miniprep plus kit (Zymo Research) or the RNeasy mini kit (Qiagen) using manufacturer's instructions. RNA samples were DNase treated, and total RNA was quantified using spectrophotometry. Integrity of RNA was checked by visualization of rRNA bands on a 1.2%

agarose gel. cDNA synthesis was performed using the iScript cDNA synthesis kit (Bio-Rad) with ~1 μg RNA as input. The Peking soybean accession used in these studies was PI 548402.

Vector construction

The *G. soja* α-SNAP (*Glyma.18G022500*), NSF (*Glyma.07G195900*), *Rhg4* (*Glyma.08G11490*), α-$SNAP_{Ch11}$ (*Glyma.11G234500)* and 5' NSF open reading frame (ORF) was amplified from generated cDNA (for α-$SNAP_{Rhg1}$, NSF and *Rhg4*) or genomic DNA (for α-$SNAP_{Ch11}$ and 5' NSF) using Kapa HiFi polymerase (Roche) and placed directly under the control of the soybean ubiquitin promoter and nopaline synthase terminator in pBlueScript using Gibson assembly (Gibson et al., 2009), and subsequently sequence-verified. For the α-$SNAP_{Rhg1}$ and the NSF, the promoter-ORF-terminator expression cassette was digested with XbaI and PstI or NotI and SalI (New England Biosciences), respectively, and the fragments were purified using the Zymoclean large fragment DNA recovery kit (Zymo Research). Purified DNA fragments were then ligated into the pSM101 binary vector using T4 DNA ligase (New England Biosciences).

For sequencing of the bridge junction, the NSF_{Ch07} amplicon, and the last exon of α-$SNAP_{Ch11}$, the *G. soja Rhg1* bridge junction, NSF_{Ch07}, or α-$SNAP_{Ch11}$ amplicon was amplified from genomic DNA using GoTaq green (Promega) and cloned into pGEM-T Easy using ligation per manufacturers recommendations (Promega). Constructs were verified by colony PCR and diagnostic digest before being sequenced by Sanger sequencing.

Transient *Agrobacterium* expression in *Nicotiana benthamiana*

A. tumefaciens strain GV3101 (pMP90) containing the indicated expression cassettes was infiltrated using a blunt tip tuberculin syringe at an OD_{600} = 0.8 into young leaves of *N. benthamiana* plants. The GV3101 cultures were grown overnight at ~28°C in media containing 25 μg/ml kanamycin and 100 μg/ml rifampicin and induced for approximately 3 h in 10 mM MES (pH = 5.6), 10 mM $MgCl_2$ and 100 mM acetosyringone prior to leaf infiltration. *N. benthamiana* plants were grown at 25°C with a photoperiod of 16 h light at 100 $\mu E \bullet m^{-2} \bullet s^{-1}$ and 8 h dark.

Antibodies and Immunoblotting

Generation and validation of rabbit polyclonal antibodies raised against *Rhg1* α-$SNAP_{LC}$, *Rhg1* α-$SNAP_{HC}$ and α-$SNAP_{WT}$ was previously described in Bayless *et al.*, 2016. Tissue preparation and immunoblots were performed essentially as in Bayless *et al.*, 2016; Bayless *et al.*, 2018. Bradford assays were performed on each leaf or root extract supernatant and equal amount of OD_{595} total protein were loaded onto SDS/PAGE gels. Immunoblots for α-SNAP and NSF 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 1:1000. Secondary horseradish peroxidase conjugated goat anti-rabbit IgG (Sigma) was added at 1:10,000 and incubated for 1 hour at room temperature on a platform shaker followed by five washes with TBS-T. Chemiluminescence signal detection was performed with SuperSignal West Dura chemiluminescent substrate (Thermo Scientific) and developed using a Chemi Doc MP chemiluminescent imager (Bio-Rad).

For antibody sensitivity assays, purified recombinant α-SNAP was serially 5-fold diluted to concentrations of 800 pg, 160 pg, and 32 pg. The proteins were then loaded onto an SDS/PAGE gel and immunoblots performed as above. To confirm loading of gel lanes, the gel was stained using ProteoSilver Kit (Sigma) according to manufacturer's instruction.

Recombinant Proteins

A vector encoding the ORF of the *G. soja* ortholog of *Glyma.18G02250* was generated and recombinant protein expression and purification were performed as in Bayless *et al.*, 2016. In brief, the ORF was cloned into the pRham-N-His-SUMO expression vector, protein was purified using PerfectPro Ni-NTA resin as per manufacturer's instruction, His and SUMO tags were cleaved by incubating the dialyzed protein with 1 unit per 100 μg protein of SUMO express protease (Lucigen) and removed by rebinding to the Ni-NTA column, and the recombinant protein was eluted in SNAP freeze-down buffer (50 mM Tris, 50 mM NaCl, 10% (w/v) glycerol, 0.5 mM TCEP, pH = 8). Both purity and quantity of protein were evaluated by Coomassie blue staining relative to BSA standards.

In vitro binding assays

In vitro binding assays were performed as previously described (Bayless et al., 2016). In brief, 20 μg of the specified recombinant α-SNAP was added to the bottom of a 1.5 mL polypropylene tube, unbound α-SNAP was removed with wash buffer, and 20 μg of either NSF or NSF_{RAN07} in NSF binding buffer (20 mM HEPES, 2 mM EDTA, 100mM KCl, 500 μM ATP, 1 mM DTT, 1% (w/v) PEG 4000) was added and incubated on ice for 10 minutes. The solution was then removed and

excess NSF was removed by washing twice with NSF binding buffer. Samples were subsequently boiled in 1X SDS loading buffer and separated on a 10% Bis-Tris SDS-PAGE gel. To detect bound NSF, the gel was stained using the ProteoSilver Kit (Sigma) according to manufacturer's instruction and quantified relative to α-SNAP abundance by densitometric analysis using ImageJ.

Nematode Resistance Tests

The nematode resistance tests were performed according to Niblack *et al.* 2009 with modifications described by Yu *et al.*, 2016. In brief, the genotypes LD10-10198 and PI 507623 were crossed in a greenhouse. F_1 seed were generated and grown to maturity to produce an F_2 population. Ninety-six F_2 plants from the cross were evaluated in a resistance bioassay.

In the first SCN bioassay, seed of the SCN indicator lines (Niblack et al. 2009), the *G. soja* accession, and the susceptible check 'Lee 74' were germinated on germination paper. After three days, individual seedlings were transplanted into PVC tubes filled with steam sterilized sandy soil and packed in plastic crocks. Each tube was infested with approximately 2000 SCN eggs derived from either HG type 0 or HG type 2.5.7 nematode populations. The experimental unit was a tube with an individual plant and the tubes were randomized using a completely randomized design. The test with each isolate was done separately and there were three replicates of the indicator lines and six replicates of the PI. After 30 days, cysts were collected by washing roots over a 250 μM sieve. The cysts were then counted, and the female index (FI) was determined using the following: FI (%) = (number of female cysts on an entry) / (average number of female cysts on susceptible control 'Lee 74') X 100. The F_2 population was evaluated

for resistance to a HG type 2.5.7 population according to the procedures listed above. In this test, each F_2 plant was placed in a separate tube and ten plants of the susceptible control Lee 74 was included. DNA was isolated from a leaf sample taken from each F_2 plant according to Yu *et al*. 2016 and tested with the genetic marker Satt309 according to Kim *et al*. (2010). A single factor analysis of variance was completed to test for an association between the genetic marker at FI.

List of primers used in the study

5’ Rhg1 Fw: TGACTAAGCAGCAAGCACAGA

5’ Rhg1 Rev: TCTTTCCGTGTACGGTCGTC

3’ Rhg1 Fw: AGAATACGTGGAGGCACAGC

3’ Rhg1Rev: CCATGCGTTACGATGCGATG

Bridge Fw-1: TTTAGCCTGCTCCTCACAAATTC

Bridge Rev-1: TTGGAGAATATGCTCTCGGTTGT

Bridge Fw-2: TTTGGGCCTTCCTCGAACA

Bridge Rev-2: CAGTGCATCAAGAGCATGCAA

5’ RAC Fw: CGATAAAGCCGCCAATTGCT

5’ RAC Rev: TCACCAAAGCTCTTGCACCT

3’ RAC Fw: GGAGATGGCCACCGAATGAA

3’ RAC Rev: GTATGTCGCTCCAGCCTTGT

RAC across Fw: CAATTGCTTCAAGCTCGCCA

RAC across Rev: GTATGTCGCTCCAGCCTTGT

RAC Gibson Fw: TTGTTGACTCGACAGAAAGCCGCCAATTGCTTCAA

RAC Gibson Rev: GGTCGAATTCGCCCTTTTCAAGCAATGTGCAGCATCGACA

a-SNAP$_{Ch11}$ Intron Gibson Fw: TTGTTGACTCGACAGGAGATAGCTCGCCAATCCC

a-SNAP$_{Ch11}$ Intron Gibson Rev: GGTCGAATTCGCCCTTTTCACTTCAATCTGGCAAATATTTCCAT

a-SNAP$_{Ch11}$ last Exon Fw: GGGGAGGCAATAGCTTAC

a-SNAP$_{Ch11}$ last Exon Rev: CAAGATATAATTCAATTCATAGTGTAGAT

Rhg4 cDNA Fw: ATGGATCCAGTAAGCGTGTGG

Rhg4 cDNA Rev: CTAATCCTTGTACTTCATTTCAG

Bridge qRT Fw: GCCGTGACTTCTTAACAAATGCAGC

Bridge qRT Rev: TGGGTAGTTTTGTTTCTTGCTCCAC

2620 qRT Fw: AAGCCCAACAGGCCAAAGAGAG

2620 qRT Rev: ACACCAAATGGGTTCGCACTTC

a-SNAP$_{Rhg1}$ Last Intron Fw: CAAGGAATTTGATAGTATGACCCCTC

a-SNAP$_{Rhg1}$ 3’ UTR Rev: AGATAAGATCAGACTCCAGCAACCTC

a-SNAP$_{Rhg1}$-Gsm pBS Gisbon Fw: TTGTTGACTCGACAGATGGCCGATCAGTTATCGAAGGG

a-SNAP$_{Rhg1}$-Gsm pBS Gibson Rev:

AGCTGGGTCGAATTCGCCCTTTTCATCAAGTAATAACCTCATGCTCCTCAAGTTC

a-SNAP$_{Rhg1}$-Gsm pRham Gibson Fw:

CGCGAACAGATTGGAGGTGCCGATCAGTTATCGAAGGGAGAGG

a-SNAP$_{Rhg1}$-Gsm pRham Gibson Rev:

GCTCAGCGGTGGCGGCCGCTCTATTAAGTAATAACCTCATGCTCCTCAAGTTC

NSF$_{Ch07}$ Detect Fw: CAACACGCCCGCGAGCGAC

NSF$_{RAN07}$ Detect Fw: CTACACGCCCGCGAACGAC

NSF Detect Rev: CTCACTTGTACGGAATCACCGG

NSF_{Ch13} Fw: GAAACAGAGAAACATAGAGGCCAT

NSF_{Ch13} Rev: CATTCAACGCAATCTGGCCG

NSF_{Ch07} 5’ Genomic Gibson Fw: TTGTTGACTCGACAGGCTCGTAAGTCGTGTTTATAGCC

NSF_{Ch07} 5’ Genomic Gibson Rev: GGTCGAATTCGCCCTTTTCAAGCCAGTAGAACAGCATCAAT

NSF_{Ch07} 3’ Genomic Fw: CTGCTGCAGAGGCGTTGAATG

NSF_{Ch07} 3’ Genomic Rev: ACCTCAATACGAGACGTACGTTTTGAT

NSF_{RAN07} cDNA pBS Gibson Fw: TTGTTGACTCGACAGATGGCGAGTCAGTTCGGG

NSF_{RAN07} cDNA pBS Gibson Rev: GGTCGAATTCGCCCTTTTCATTATAACCTAACAACATCCTGGAGGC

4.6 Acknowledgments

We thank Adam Bayless for early contributions to the study and Alison Colgrove for assistance with nematode experiments. Funding for this study was provided by United Soybean Board awards to A.B. and B.D., by the Department of Plant Pathology and the Agricultural Experiment Station Hatch program at University of Wisconsin - Madison, and by National Science Foundation Predoctoral Fellowship and NIH Molecular Biosciences Training Grant awards to D.G.

4.7 References

Arelli, P. R., Sleper, D. A., Yue, P., & Wilcox, J. A. (2000). Soybean Reaction to Races 1 and 2 of *Heterodera glycines*. *Crop Science*, *40*(3), 824–826. https://doi.org/10.2135/cropsci2000.403824x

Barnard, R. J. O., Morgan, A., & Burgoyne, R. D. (1997). Stimulation of NSF ATPase Activity by α-SNAP Is Required for SNARE Complex Disassembly and Exocytosis. *Journal of Cell Biology*, *139*(4), 875–883. https://doi.org/10.1083/jcb.139.4.875

Barszczewski, M., Chua, J. J., Stein, A., Winter, U., Heintzmann, R., Zilly, F. E., … Jahn, R. (2008). A Novel Site of Action for α-SNAP in the SNARE Conformational Cycle Controlling Membrane Fusion. *Molecular Biology of the Cell*, *19*(3), 776–784. https://doi.org/10.1091/mbc.e07-05-0498

Bayless, A. M., Smith, J. M., Song, J., McMinn, P. H., Teillet, A., August, B. K., & Bent, A. F. (2016). Disease resistance through impairment of α-SNAP–NSF interaction and vesicular trafficking by soybean *Rhg1*. *Proceedings of the National Academy of Sciences*, *113*(47), E7375–E7382. https://doi.org/10.1073/pnas.1610150113

Bayless, A. M., Zapotocny, R. W., Grunwald, D. J., Amundson, K. K., Diers, B. W., & Bent, A. F. (2018). An atypical N-ethylmaleimide sensitive factor enables the viability of nematode-resistant *Rhg1* soybeans. *Proceedings of the National Academy of Sciences*, *115*(19), E4512–E4521. https://doi.org/10.1073/pnas.1717070115

Bayless, A. M., Zapotocny, R. W., Han, S., Grunwald, D. J., Amundson, K. K., & Bent, A. F. (2019). The *rhg1-a* (*Rhg1* low-copy) nematode resistance source harbors a copia-family retrotransposon within the *Rhg1*-encoded α-SNAP gene. *Plant Direct*, *3*(8), e00164. https://doi.org/10.1002/pld3.164

Bekal, S., Domier, L. L., Gonfa, B., Lakhssassi, N., Meksem, K., & Lambert, K. N. (2015). A SNARE-Like Protein and Biotin Are Implicated in Soybean Cyst Nematode Virulence. *PLOS ONE*, *10*(12), e0145601. https://doi.org/10.1371/journal.pone.0145601

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*(1), 44–49. https://doi.org/10.1007/s00122-005-1970-3

Brzostowski, L. F., & Diers, B. W. (2017). Pyramiding of Alleles from Multiple Sources Increases the Resistance of Soybean to Highly Virulent Soybean Cyst Nematode Isolates. *Crop Science*, *57*(6), 2932–2941. https://doi.org/10.2135/cropsci2016.12.1007

Colgrove, A., & Niblack, T. (2008). Correlation of female indices from virulence assays on inbred lines and field populations of *Heterodera glycines*. *The Journal of Nematology*, *40*(1), 39–45.

Concibido, V. C., Diers, B. W., & Arelli, P. R. (2004). A Decade of QTL Mapping for Cyst Nematode Resistance in Soybean. *Crop Science*, *44*(4), 1121–1131. https://doi.org/10.2135/cropsci2004.1121

Concibido, V., La Vallee, B., Mclaird, P., Pineda, N., Meyer, J., Hummel, L., … Delannay, X. (2003). Introgression of a quantitative trait locus for yield from *Glycine soja* into commercial

soybean cultivars. *Theoretical and Applied Genetics*, *106*(4), 575–582. https://doi.org/10.1007/s00122-002-1071-5

Cook, D. E., Lee, T. G., Guo, X., Melito, S., Wang, K., Bayless, A. M., … Bent, A. F. (2012). Copy Number Variation of Multiple Genes at *Rhg1* Mediates Nematode Resistance in Soybean. *Science*, *338*(6111), 1206–1209. https://doi.org/10.1126/science.1228746

Cook, David E., Bayless, A. M., Wang, K., Guo, X., Song, Q., Jiang, J., & Bent, A. F. (2014). Distinct Copy Number, Coding Sequence, and Locus Methylation Patterns Underlie *Rhg1*-Mediated Soybean Resistance to Soybean Cyst Nematode. *Plant Physiology*, *165*(2), 630–647. https://doi.org/10.1104/pp.114.235952

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics (Oxford, England)*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Davis, E. L., Hussey, R. S., Mitchum, M. G., & Baum, T. J. (2008). Parasitism proteins in nematode–plant interactions. *Current Opinion in Plant Biology*, *11*(4), 360–366. https://doi.org/10.1016/j.pbi.2008.04.003

Donald, P., Pierson, P., St. Martin, S., Sellers, P., Noel, G., MacGuidwin, A., … Wyson, D. (2006). Accessing *Heterodera glycines* - Resistant and Susceptible Cultivar Yield Response. *The Journal of Nematology*, *38*(1), 76–82.

Dong, J., Zielinski, R. E., & Hudson, M. E. (2020). t-SNAREs Bind the *Rhg1* α-SNAP and Mediate Soybean Cyst Nematode Resistance. *The Plant Journal*. https://doi.org/10.1111/tpj.14923

Gardner, M., Dhroso, A., Johnson, N., Davis, E. L., Baum, T. J., Korkin, D., & Mitchum, M. G. (2018). Novel global effector mining from the transcriptome of early life stages of the soybean cyst nematode *Heterodera glycines*. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-20536-5

Gheysen, G., & Mitchum, M. G. (2011). How nematodes manipulate plant development pathways for infection. *Current Opinion in Plant Biology*, *14*(4), 415–421. https://doi.org/10.1016/j.pbi.2011.03.012

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, *6*(5), 343–345. https://doi.org/10.1038/nmeth.1318

Graham, P. H., & Vance, C. P. (2003). Legumes: Importance and Constraints to Greater Use. *Plant Physiology*, *131*(3), 872–877. https://doi.org/10.1104/pp.017004

Germplasm Resources Information Network [Internet]. Beltsville (MD): United States Department of Agriculture, Agricultural Research Service. 3 January 2021. Available from: http://www.ars-grin.gov/.

Hesler, L. S. (2013). Resistance to soybean aphid among wild soybean lines under controlled conditions. *Crop Protection*, *53*, 139–146. https://doi.org/10.1016/j.cropro.2013.06.016

Hoefle, C., & Hückelhoven, R. (2008). Enemy at the gates: traffic at the plant cell pathogen interface. *Cellular Microbiology*, *10*(12), 2400–2407. https://doi.org/10.1111/j.1462-5822.2008.01238.x

Huson, D. H., & Bryant, D. (2005). Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution*, *23*(2), 254–267. https://doi.org/10.1093/molbev/msj030

Hyten, D. L., Song, Q., Zhu, Y., Choi, I.-Y., Nelson, R. L., Costa, J. M., … Cregan, P. B. (2006). Impacts of genetic bottlenecks on soybean genome diversity. *Proceedings of the National Academy of Sciences*, *103*(45), 16666–16671. https://doi.org/10.1073/pnas.0604379103

Inada, N., & Ueda, T. (2014). Membrane Trafficking Pathways and their Roles in Plant–Microbe Interactions. *Plant and Cell Physiology*, *55*(4), 672–686. https://doi.org/10.1093/pcp/pcu046

Jahn, R., & Scheller, R. H. (2006). SNAREs — engines for membrane fusion. *Nature Reviews Molecular Cell Biology*, *7*(9), 631–643. https://doi.org/10.1038/nrm2002

Kabelka, E. A., Carlson, S. R., & Diers, B. W. (2005). Localization of Two Loci that Confer Resistance to Soybean Cyst Nematode from *Glycine soja* PI 468916. *Crop Science*, *45*(6), 2473–2481. https://doi.org/10.2135/cropsci2005.0027

Kabelka, E. A., Carlson, S. R., & Diers, B. W. (2006). *Glycine soja* PI 468916 SCN Resistance Loci's Associated Effects on Soybean Seed Yield and Other Agronomic Traits. *Crop Science*, *46*(2), 622–629. https://doi.org/10.2135/cropsci2005.06-0131

Kandoth, P. K., Ithal, N., Recknor, J., Maier, T., Nettleton, D., Baum, T. J., & Mitchum, M. G. (2011). The Soybean *Rhg1* Locus for Resistance to the Soybean Cyst Nematode *Heterodera glycines* Regulates the Expression of a Large Number of Stress- and Defense-Related Genes in Degenerating Feeding Cells. *Plant Physiology*, *155*(4), 1960–1975. https://doi.org/10.1104/pp.110.167536

Khan, R., Alkharouf, N., MacDonald, M., Chouikha, I., Meyer, S., Grefenstette, J., … Matthews, B. (2004). Microarray Analysis of Gene Expression in Soybean Roots Susceptible to the

Soybean Cyst Nematode Two Days Post Invasion. *The Journal of Nematology*, *36*(3), 214–248.

Kim, M., & Diers, B. W. (2013). Fine Mapping of the SCN Resistance QTL and from PI 468916. *Crop Science*, *53*(3), 775. https://doi.org/10.2135/cropsci2012.07.0425

Kim, M., Hyten, D. L., Bent, A. F., & Diers, B. W. (2010). Fine Mapping of the SCN Resistance Locus *rhg1-b* from PI 88788. *The Plant Genome Journal*, *3*(2), 81. https://doi.org/10.3835/plantgenome2010.02.0001

Klink, V. P., Alkharouf, N., MacDonald, M., & Matthews, B. (2005). Laser Capture Microdissection (LCM) and Expression Analyses of *Glycine max* (Soybean) Syncytium Containing Root Regions Formed by the Plant Pathogen *Heterodera glycines* (Soybean Cyst Nematode). *Plant Molecular Biology*, *59*(6), 965–979. https://doi.org/10.1007/s11103-005-2416-7

Klink, V. P., Hosseini, P., Matsye, P., Alkharouf, N. W., & Matthews, B. F. (2009). A gene expression analysis of syncytia laser microdissected from the roots of the *Glycine max* (soybean) genotype PI 548402 (Peking) undergoing a resistant reaction after infection by *Heterodera glycines* (soybean cyst nematode). *Plant Molecular Biology*, *71*(6), 525–567. https://doi.org/10.1007/s11103-009-9539-1

Klink, V. P., Hosseini, P., Matsye, P. D., Alkharouf, N. W., & Matthews, B. F. (2010). Syncytium gene expression in *Glycine max*[PI 88788] roots undergoing a resistant reaction to the parasitic nematode *Heterodera glycines*. *Plant Physiology and Biochemistry*, *48*(2–3), 176–193. https://doi.org/10.1016/j.plaphy.2009.12.003

Klink, V. P., Overall, C. C., Alkharouf, N. W., MacDonald, M. H., & Matthews, B. F. (2007a). A time-course comparative microarray analysis of an incompatible and compatible response by *Glycine max* (soybean) to *Heterodera glycines* (soybean cyst nematode) infection. *Planta*, *226*(6), 1423–1447. https://doi.org/10.1007/s00425-007-0581-4

Klink, V. P., Overall, C. C., Alkharouf, N. W., MacDonald, M. H., & Matthews, B. F. (2007b). 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*(6), 1389–1409. https://doi.org/10.1007/s00425-007-0578-z

Koenning, S. R., & Wrather, J. A. (2010). Suppression of Soybean Yield Potential in the Continental United States by Plant Diseases from 2006 to 2009. *Plant Health Progress*, *11*(1), 5. https://doi.org/10.1094/php-2010-1122-01-rs

Kopisch-Obuch, F. J., & Diers, B. W. (2005). Segregation at the SCN resistance locus *rhg1* in soybean is distorted by an association between the resistance allele and reduced field

emergence. *Theoretical and Applied Genetics*, *112*(2), 199–207. https://doi.org/10.1007/s00122-005-0104-2

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Kyndt, T., Vieira, P., Gheysen, G., & de Almeida-Engler, J. (2013). Nematode feeding sites: unique organs in plant roots. *Planta*, *238*(5), 807–818. https://doi.org/10.1007/s00425-013-1923-z

Lakhssassi, N., Liu, S., Bekal, S., Zhou, Z., Colantonio, V., Lambert, K., … Meksem, K. (2017). Characterization of the Soluble NSF Attachment Protein gene family identifies two members involved in additive resistance to a plant pathogen. *Scientific Reports*, *7*(1). https://doi.org/10.1038/srep45226

Lakhssassi, N., Piya, S., Bekal, S., Liu, S., Zhou, Z., Bergounioux, C., … Meksem, K. (2020). A pathogenesis-related protein GmPR08-Bet VI promotes a molecular interaction between the GmSHMT08 and GmSNAP18 in resistance to *Heterodera glycines*. *Plant Biotechnology Journal*, *18*(8), 1810–1829. https://doi.org/10.1111/pbi.13343

Lee, J. S., Yoo, M., Jung, J. K., Bilyeu, K. D., Lee, J.-D., & Kang, S. (2015). Detection of novel QTLs for foxglove aphid resistance in soybean. *Theoretical and Applied Genetics*, *128*(8), 1481–1488. https://doi.org/10.1007/s00122-015-2519-8

Lee, J.-D., Shannon, J. G., Vuong, T. D., & Nguyen, H. T. (2009). Inheritance of Salt Tolerance in Wild Soybean (*Glycine soja* Sieb. and Zucc.) Accession PI483463. *Journal of Heredity*, *100*(6), 798–801. https://doi.org/10.1093/jhered/esp027

Lee, T. G., Diers, B. W., & Hudson, M. E. (2016). An efficient method for measuring copy number variation applied to improvement of nematode resistance in soybean. *The Plant Journal*, *88*(1), 143–153. https://doi.org/10.1111/tpj.13240

Lee, T. G., Kumar, I., Diers, B. W., & Hudson, M. E. (2015). Evolution and selection of *Rhg1*, a copy-number variant nematode-resistance locus. *Molecular Ecology*, *24*(8), 1774–1791. https://doi.org/10.1111/mec.13138

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*, *26*(5), 589–595. https://doi.org/10.1093/bioinformatics/btp698

Liu, S., Kandoth, P. K., Lakhssassi, N., Kang, J., Colantonio, V., Heinz, R., … Meksem, K. (2017). The soybean GmSNAP18 gene underlies two types of resistance to soybean cyst nematode. *Nature Communications*, *8*(1). https://doi.org/10.1038/ncomms14822

Liu, S., Kandoth, P. K., Warren, S. D., Yeckel, G., Heinz, R., Alden, J., … Meksem, K. (2012). A soybean cyst nematode resistance gene points to a new mechanism of plant resistance to pathogens. *Nature*, *492*(7428), 256–260. https://doi.org/10.1038/nature11651

Liu, Y., Du, H., Li, P., Shen, Y., Peng, H., Liu, S., … Tian, Z. (2020). Pan-Genome of Wild and Cultivated Soybeans. *Cell*, *182*(1), 162-176.e13. https://doi.org/10.1016/j.cell.2020.05.023

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10. https://doi.org/10.14806/ej.17.1.200

Marz, K. E., Lauer, J. M., & Hanson, P. I. (2003). Defining the SNARE Complex Binding Surface of α-SNAP. *Journal of Biological Chemistry*, *278*(29), 27000–27008. https://doi.org/10.1074/jbc.m302003200

Matsye, P. D., Lawrence, G. W., Youssef, R. M., Kim, K.-H., Lawrence, K. S., Matthews, B. F., & Klink, V. P. (2012). The expression of a naturally occurring, truncated allele of an α-SNAP gene suppresses plant parasitic nematode infection. *Plant Molecular Biology*, *80*(2), 131–155. https://doi.org/10.1007/s11103-012-9932-z

McCarville, M. T., Marett, C. C., Mullaney, M. P., Gebhart, G. D., & Tylka, G. L. (2017). Increase in Soybean Cyst Nematode Virulence and Reproduction on Resistant Soybean Varieties in Iowa From 2001 to 2015 and the Effects on Soybean Yields. *Plant Health Progress*, *18*(3), 146–155. https://doi.org/10.1094/php-rs-16-0062

Meksem, K., Pantazopoulos, P., Njiti, V. N., Hyten, L. D., Arelli, P. R., & Lightfoot, D. A. (2001). 'Forrest' resistance to the soybean cyst nematode is bigenic: saturation mapping of the *Rhg1* and *Rhg4* loci. *Theoretical and Applied Genetics*, *103*(5), 710–717. https://doi.org/10.1007/s001220100597

Naydenov, N. G., Harris, G., Brown, B., Schaefer, K. L., Das, S. K., Fisher, P. B., & Ivanov, A. I. (2011). Loss of Soluble N-ethylmaleimide-sensitive Factor Attachment Protein α (αSNAP) Induces Epithelial Cell Apoptosis via Down-regulation of Bcl-2 Expression and Disruption of the Golgi. *Journal of Biological Chemistry*, *287*(8), 5928–5941. https://doi.org/10.1074/jbc.m111.278358

Niblack, T. L., Colgrove, A. L., Colgrove, K., & Bond, J. P. (2008). Shift in Virulence of Soybean Cyst Nematode is Associated with Use of Resistance from PI 88788. *Plant Health Progress*, *9*(1), 29. https://doi.org/10.1094/php-2008-0118-01-rs

Niblack, T. L., Lambert, K. N., & Tylka, G. L. (2006). A Model Plant Pathogen from the Kingdom Animalia: *Heterodera glycines*, the Soybean Cyst Nematode. *Annual Review of Phytopathology*, *44*(1), 283–303. https://doi.org/10.1146/annurev.phyto.43.040204.140218

Niblack, T. L., Tylka, G. L., Arelli, P., Bond, J., Diers, B., Donald, P., … Wilcox, J. (2009). A Standard Greenhouse Method for Assessing Soybean Cyst Nematode Resistance in Soybean: SCE08 (Standardized Cyst Evaluation 2008). *Plant Health Progress*, *10*(1), 33. https://doi.org/10.1094/php-2009-0513-01-rv

Patil, G. B., Lakhssassi, N., Wan, J., Song, L., Zhou, Z., Klepadlo, M., … Nguyen, H. T. (2019). Whole-genome re-sequencing reveals the impact of the interaction of copy number variants of the *rhg1* and *Rhg4* genes on broad-based resistance to soybean cyst nematode. *Plant Biotechnology Journal*, *17*(8), 1595–1611. https://doi.org/10.1111/pbi.13086

Poland, J. A., Brown, P. J., Sorrells, M. E., & Jannink, J.-L. (2012). Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. *PLoS ONE*, *7*(2), e32253. https://doi.org/10.1371/journal.pone.0032253

Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Van der Auwera, G. A., … Banks, E. (2017). Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv*. https://doi.org/10.1101/201178

Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., … Jackson, S. A. (2010). Genome sequence of the palaeopolyploid soybean. *Nature*, *463*(7278), 178–183. https://doi.org/10.1038/nature08670

Shi, Z., Liu, S., Noe, J., Arelli, P., Meksem, K., & Li, Z. (2015). SNP identification and marker assay development for high-throughput selection of soybean cyst nematode resistance. *BMC Genomics*, *16*(1). https://doi.org/10.1186/s12864-015-1531-3

Song, Q., Hyten, D. L., Jia, G., Quigley, C. V., Fickus, E. W., Nelson, R. L., & Cregan, P. B. (2013). Development and Evaluation of SoySNP50K, a High-Density Genotyping Array for Soybean. *PLoS ONE*, *8*(1), e54985. https://doi.org/10.1371/journal.pone.0054985

Song, Q., Hyten, D. L., Jia, G., Quigley, C. V., Fickus, E. W., Nelson, R. L., & Cregan, P. B. (2015). Fingerprinting Soybean Germplasm and Its Utility in Genomic Research. *G3: Genes|Genomes|Genetics*, *5*(10), 1999–2006. https://doi.org/10.1534/g3.115.019000

Soystats.org. http://soystats.com/. (n.d.). Retrieved August 2, 2020, from http://soystats.com

Stacey, G. (Ed. (2010). *Genetics and genomics of soybean.* New York: Springer.

Stecher, G., Tamura, K., & Kumar, S. (2020). Molecular Evolutionary Genetics Analysis (MEGA) for macOS. *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msz312

Torkamaneh, D., Laroche, J., Valliyodan, B., O'Donoughue, L., Cober, E., Rajcan, I., … Belzile, F. (2019). Soybean Haplotype Map (GmHapMap): A Universal Resource for Soybean Translational and Functional Genomics. *BioRxiv*. https://doi.org/10.1101/534578

Tran, D. T., Steketee, C. J., Boehm, J. D., Noe, J., & Li, Z. (2019). Genome-Wide Association Analysis Pinpoints Additional Major Genomic Regions Conferring Resistance to Soybean Cyst Nematode (*Heterodera glycines* Ichinohe). *Frontiers in Plant Science*, *10*. https://doi.org/10.3389/fpls.2019.00401

Tuberosa, R. (2012). Phenotyping for drought tolerance of crops in the genomics era. *Frontiers in Physiology*, *3*. https://doi.org/10.3389/fphys.2012.00347

Usovsky, M., Ye, H., Vuong, T. D., Patil, G. B., Wan, J., Zhou, L., & Nguyen, H. T. (2020). Fine-mapping and characterization of qSCN18, a novel QTL controlling soybean cyst nematode resistance in PI 567516C. *Theoretical and Applied Genetics*. https://doi.org/10.1007/s00122-020-03718-6

Varley, K. E., & Mitra, R. D. (2010). Bisulfite Patch PCR enables multiplexed sequencing of promoter methylation across cancer samples. *Genome Research*, *20*(9), 1279–1287. https://doi.org/10.1101/gr.101212.109

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

Wang, D., Diers, B. W., Arelli, P. R., & Shoemaker, R. C. (2001). Loci underlying resistance to Race 3 of soybean cyst nematode in *Glycine soja* plant introduction 468916. *Theoretical and Applied Genetics*, *103*(4), 561–566. https://doi.org/10.1007/pl00002910

Webb, D. M., Baltazar, B. M., Rao-Arelli, A. P., Schupp, J., Clayton, K., Keim, P., & Beavis, W. D. (1995). Genetic mapping of soybean cyst nematode race-3 resistance loci in the soybean PI 437.654. *Theoretical and Applied Genetics*, *91*(4), 574–581. https://doi.org/10.1007/bf00223282

Winter, S. M. J., Shelp, B. J., Anderson, T. R., Welacky, T. W., & Rajcan, I. (2006). QTL associated with horizontal resistance to soybean cyst nematode in *Glycine soja* PI464925B. *Theoretical and Applied Genetics*, *114*(3), 461–472. https://doi.org/10.1007/s00122-006-0446-4

Wu, X.-Y., Zhou, G.-C., Chen, Y.-X., Wu, P., Liu, L.-W., Ma, F.-F., … Wang, B. (2016). Soybean Cyst Nematode Resistance Emerged via Artificial Selection of Duplicated Serine Hydroxymethyltransferase Genes. *Frontiers in Plant Science*, *7*. https://doi.org/10.3389/fpls.2016.00998

Xie, M., Chung, C. Y.-L., Li, M.-W., Wong, F.-L., Wang, X., Liu, A., … Lam, H.-M. (2019). A reference-grade wild soybean genome. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-09142-9

Yu, N., & Diers, B. W. (2017). Fine mapping of the SCN resistance QTL cqSCN-006 and cqSCN-007 from *Glycine soja* PI 468916. *Euphytica*, *213*(2). https://doi.org/10.1007/s10681-016-1791-2

Yu, N., Lee, T. G., Rosa, D. P., Hudson, M., & Diers, B. W. (2016). Impact of *Rhg1* copy number, type, and interaction with *Rhg4* on resistance to *Heterodera glycines* in soybean. *Theoretical and Applied Genetics*, *129*(12), 2403–2412. https://doi.org/10.1007/s00122-016-2779-y

Zhang, H., Kjemtrup-Lovelace, S., Li, C., Luo, Y., Chen, L. P., & Song, B.-H. (2017). Comparative RNA-Seq Analysis Uncovers a Complex Regulatory Network for Soybean Cyst Nematode Resistance in Wild Soybean (*Glycine soja*). *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-09945-0

Zhang, H., Li, C., Davis, E. L., Wang, J., Griffin, J. D., Kofsky, J., & Song, B.-H. (2016). Genome-Wide Association Study of Resistance to Soybean Cyst Nematode (*Heterodera glycines*) HG Type 2.5.7 in Wild Soybean (*Glycine soja*). *Frontiers in Plant Science*, *7*. https://doi.org/10.3389/fpls.2016.01214

Zhang, H., Song, Q., Griffin, J. D., & Song, B.-H. (2017). Genetic architecture of wild soybean (*Glycine soja*) response to soybean cyst nematode (*Heterodera glycines*). *Molecular Genetics and Genomics*, *292*(6), 1257–1265. https://doi.org/10.1007/s00438-017-1345-x

Zhao, C., Slevin, J. T., & Whiteheart, S. W. (2007). Cellular functions of NSF: Not just SNAPs and SNAREs. *FEBS Letters*, *581*(11), 2140–2149. https://doi.org/10.1016/j.febslet.2007.03.032

Zhao, M., Wu, S., Zhou, Q., Vivona, S., Cipriano, D. J., Cheng, Y., & Brunger, A. T. (2015). Mechanistic insights into the recycling machine of the SNARE complex. *Nature*, *518*(7537), 61–67. https://doi.org/10.1038/nature14148

Zhao, C., Slevin, J. T., & Whiteheart, S. W. (2007). Cellular functions of NSF: Not just SNAPs and SNAREs. *FEBS Letters*, *581*(11), 2140–2149. https://doi.org/10.1016/j.febslet.2007.03.032

4.8 Figures

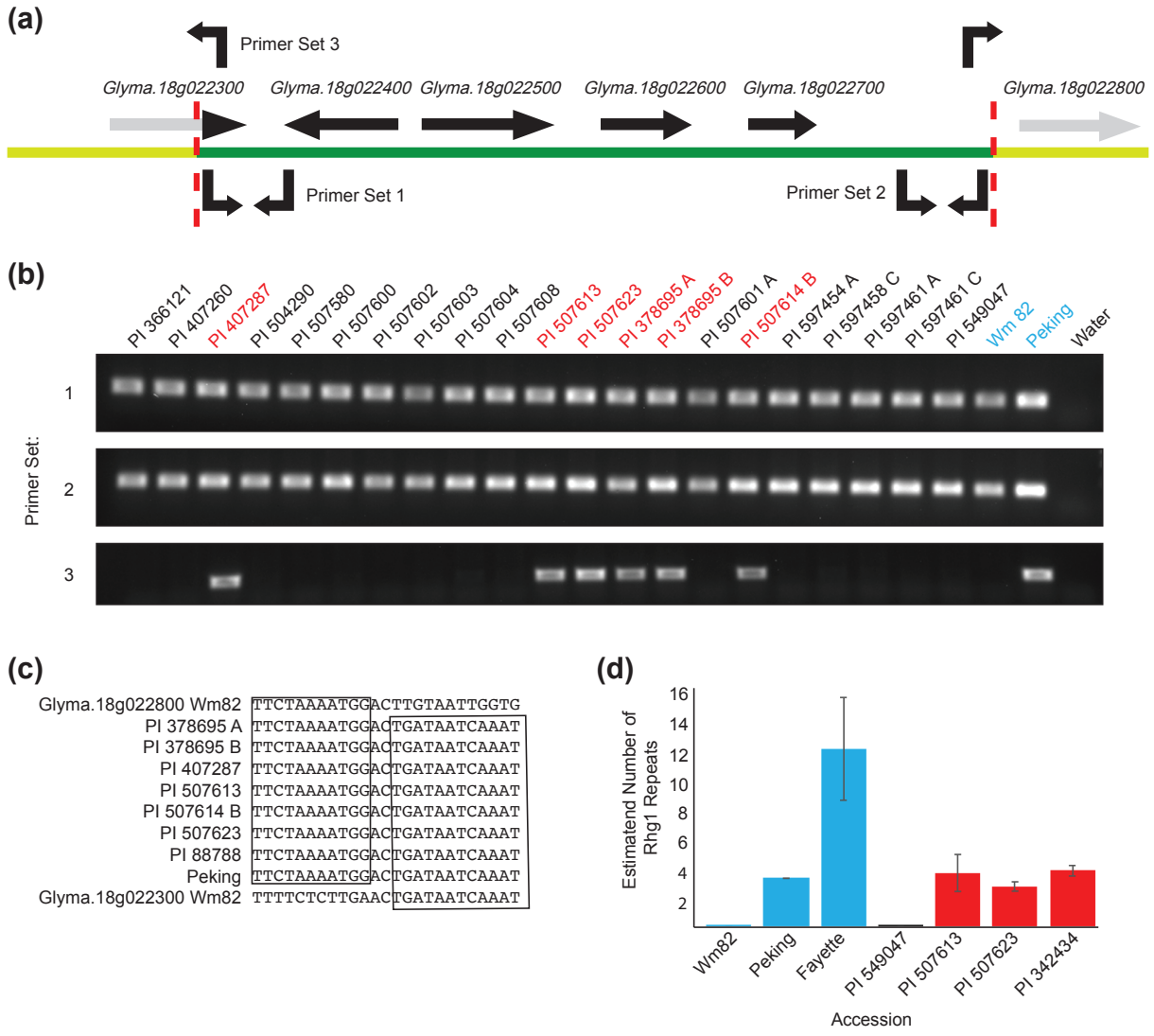


Figure 1. Multiple *Glycine soja* carry the *Resistance to Heterodera glycines 1* (*Rhg1*) repeat junction. (a) Schematic of PCR primers used in (b). (b) Results of PCR using Primer Set 3 primers that orient outwards from within the 5' and 3' ends of the ~31 kb segment of *Rhg1* found in both single- and multi-copy lines. Product indicates tandem duplication of the *Rhg1* locus. (c) DNA sequence of the *Rhg1* repeat junction found in *Glycine soja* accessions that contain a *Rhg1*

duplication event. (d) *Rhg1* copy number qPCR on *G. soja* accessions as well as domesticated soybean cultivars Wm82, Peking, and Fayette. Error bars show standard error of the mean.

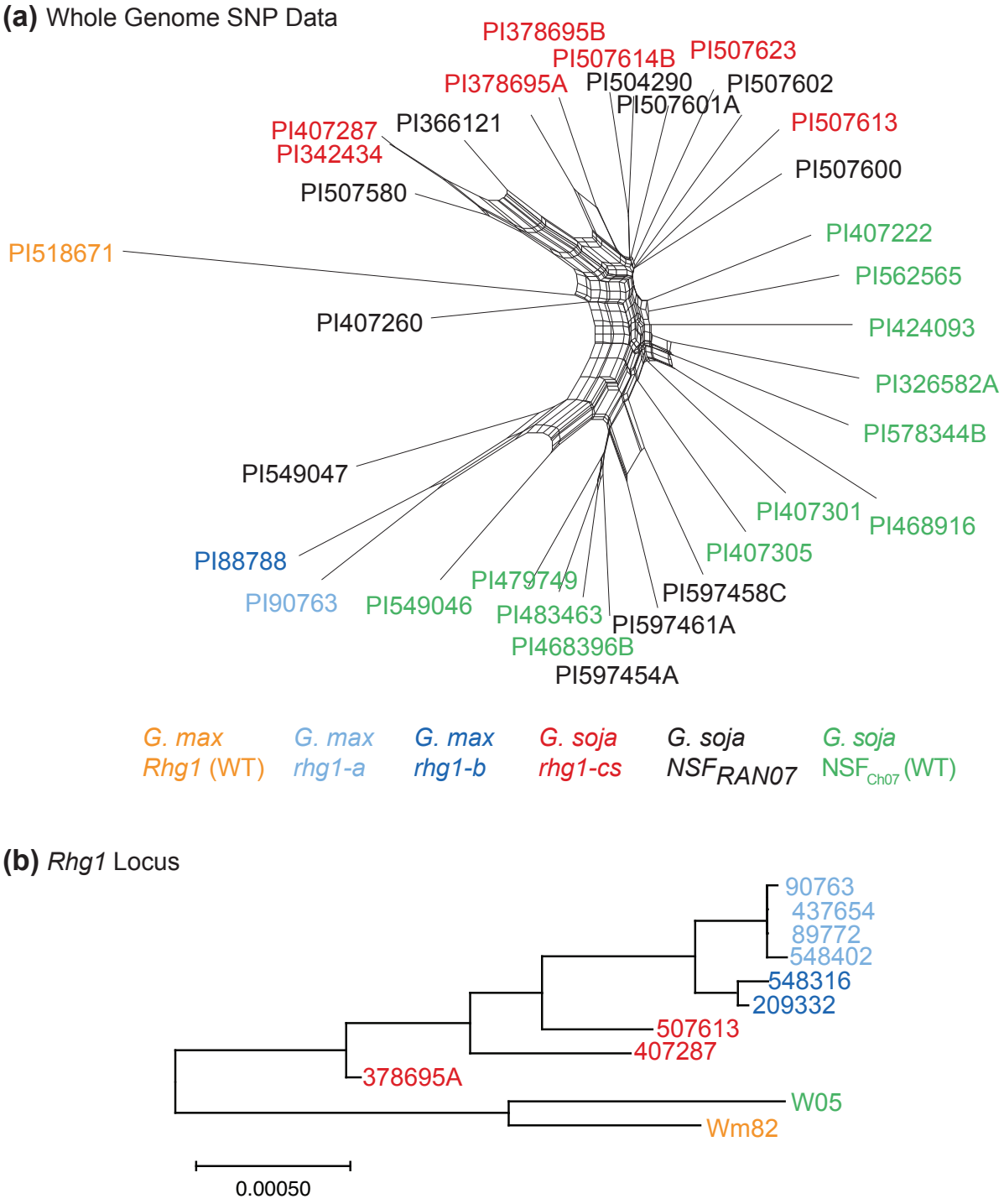


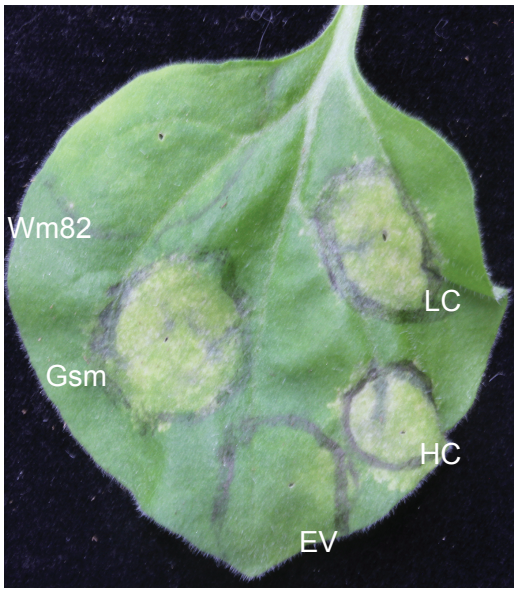
Figure 2. **Multicopy *Resistance to Heterodera glycines 1* (*Rhg1*)-containing *Glycine soja* group separately from multicopy *Rhg1*-containing *G. max* at the levels of whole genome and *Rhg1* locus.** (a) Unrooted phylogeny assessing relatedness of whole-genome SNP signatures of a

representative set of *G. soja* accessions. Accessions containing *rhg1-cs* and NSF_{RAN07} (red) were analyzed alongside those containing NSF_{RAN07} but lacking the multicopy *rhg1-cs* (black) and a representative semi-random set of non-*rhg1-cs* non- NSF_{RAN07} *G. soja* accessions (green). Note that *G. max* accessions carrying *rhg1-a* or *rhg1-b* also carry NSF_{RAN07}. (b) Rooted phylogeny generated using genomic sequences of the *Rhg1* locus from the denoted accessions, comparing relatedness of the *rhg1-a* (light blue), *rhg1-b* (dark blue) and *rhg1-cs* (red) loci as well as the *Rhg1* loci from accession W05 (a *G. soja* carrying a single-copy *Rhg1*; green) and *G. max* cv. Williams 82 (orange).

(a)

α–SNAP$_{Rhg1}$Wm82 (WT) KAKELEEDD–LT
α–SNAP$_{Rhg1}$*G. soja* (WT) KAKELEEDD–LT
α–SNAP$_{Rhg1}$HC (*rhg1–b*) KAKELEQHEAIT
α–SNAP$_{Rhg1}$Gsm(*rhg1–cs*) KAKELEEHEVIT
α–SNAP$_{Rhg1}$LC (*rhg1–a*) KAKELEEYEVIT

(b)



(c)

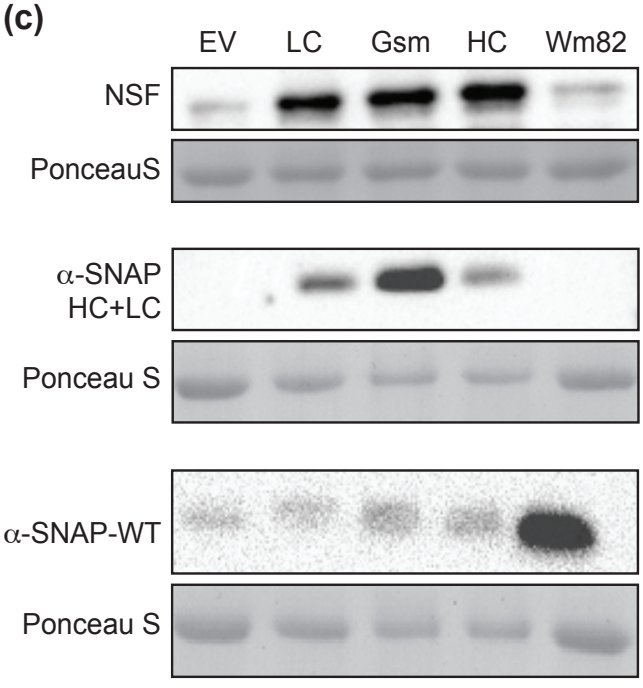


Figure 3. Unique *Glycine soja* alpha-Soluble NSF Attachment Protein (α-SNAP) induces cell death and N-ethylmaleimide Sensitive Factor (NSF) hyperaccumulation in *Nicotiana*

***benthamiana*.** (a) Alignment of C-terminal amino acids of the *Resistance to Heterodera glycines 1* (*Rhg1*) α-SNAP from multicopy *rhg1-cs Glycine soja* (α-SNAP$_{Rhg1}$Gsm) compared to *Rhg1* α-SNAPs from single-copy SCN-susceptible cv. Williams 82 (Wm82; WT) or *G. soja* W05 (*G. soja*; WT), and multicopy SCN-resistant *G. max* cvs. PI 88788 (HC; *rhg1-b*) and Peking (LC; *rhg1-a*). (b) Representative *N. benthamiana* leaf seven days after infiltration with *A. tumefaciens* expressing empty vector negative control or the denoted *Rhg1* α-SNAP with no epitope tag [abbreviations as in (a)]. (c) Immunoblots detecting levels of endogenous *N. benthamiana* NSF or α-SNAP (α-SNAP-WT), as well as recombinant *Glycine* α-SNAPs delivered via *A. tumefaciens* as in (b). Same sample was used for entire column in (c) but probed with three separate antibodies. Leaf tissue was harvested approximately 60 hours after infiltration; Ponceau S staining indicates relative levels of total protein in each sample.

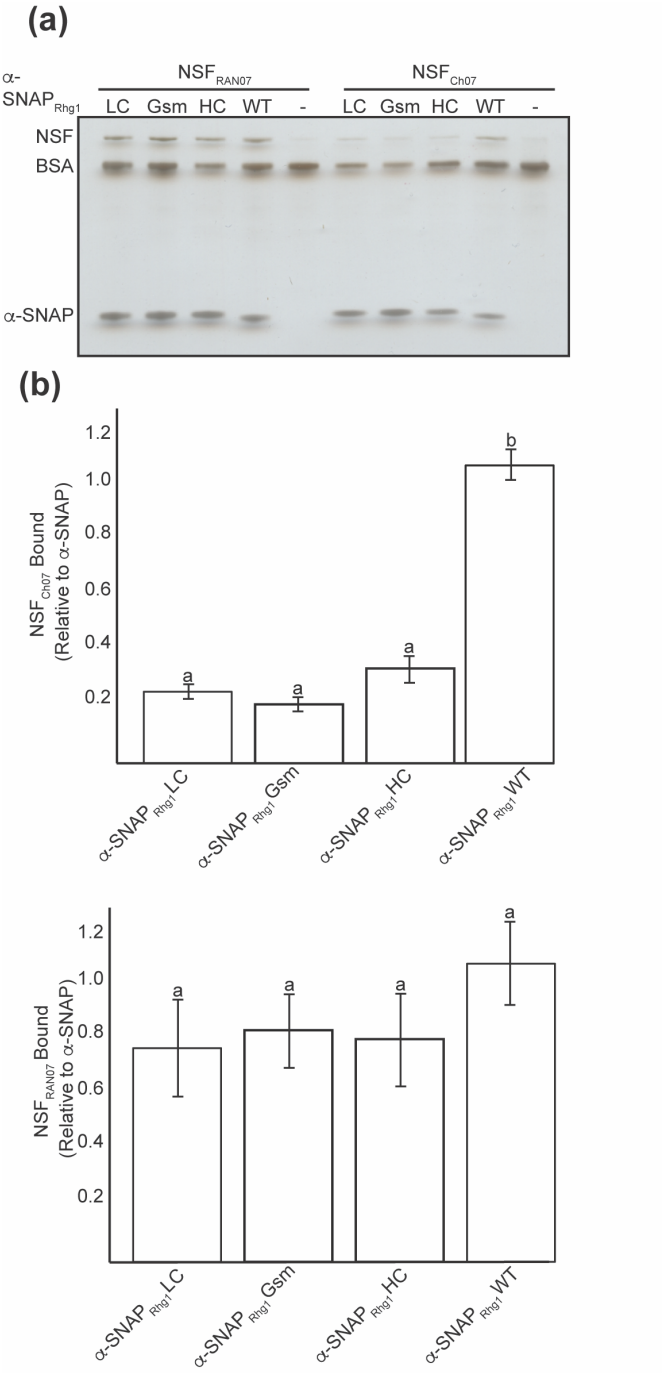


Figure 4. *Glycine soja rhg1-cs* alpha-Soluble NSF Attachment Protein (α-SNAP$_{Rhg1}$Gsm) interacts well with NSF$_{RAN07}$ N-ethylmaleimide Sensitive Factor protein but poorly with wild-type NSF. (a) Representative silver-stained SDS/PAGE gel of recombinant soybean wild-type NSF$_{Ch07}$ or NSF$_{RAN07}$ bound *in vitro* by different recombinant α-SNAP$_{Rhg1}$ proteins in relative

binding affinity assay. LC: α-SNAP$_{Rhg1}$LC; Gsm: α-SNAP$_{Rhg1}$Gsm; HC: α-SNAP$_{Rhg1}$HC; WT: Wild type (single-copy) α-SNAP$_{Rhg1}$; BSA, bovine serum albumin (carrier protein). (b) Densitometric quantification of relative amount of NSF$_{Ch07}$ (upper graph) or NSF$_{RAN07}$ (lower graph) bound by the indicated α-SNAP, collecting results from three independent experiments. Error bars show standard error of the mean. Bars with same letter are not statistically different (ANOVA, Tukey HSD $p > 0.05$).

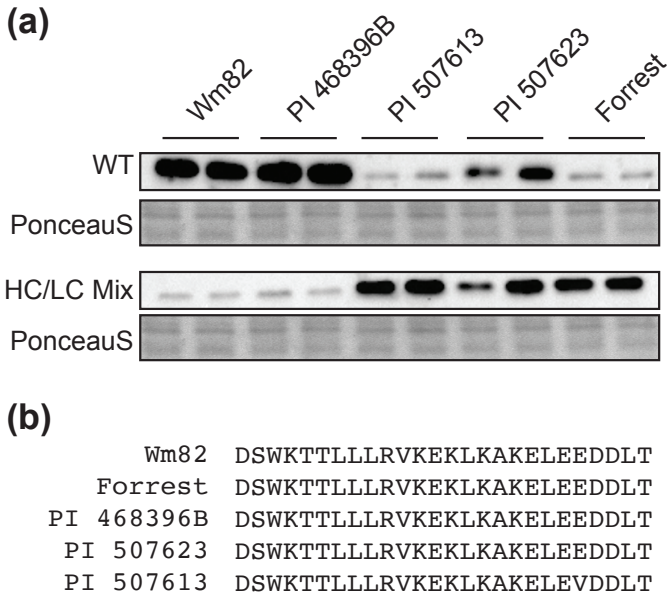


Figure 5. PI 507613 encodes a unique alpha Soluble NSF Attachment Protein on chromosome 11 (α-SNAP$_{Ch11}$). (a) Immunoblots of root protein extracts in *Resistance to Heterodera glycines* (*Rhg1*) *rhg1-cs*-containing *G. soja* and controls shows low apparent expression of α-SNAP$_{WT}$ in PI 507613. Paired lanes are replicate samples from same accession. WT and HC/LC Mix refer to the polyclonal antibody used for detection. (b) Alignment of C-terminal amino acids of α-SNAP$_{Ch11}$ in PI 507613 showing the E$_{284}$V polymorphism, relative to other *rhg1-cs*-containing *G. soja* (PI 507623), single-copy *G. soja* (PI 468396B), and *G. max* cultivars Williams 82 and Forrest.

Table 1: Relative Soybean Cyst Nematode cyst production on select *G. soja* accessions and HG type indicator *G. max* lines, for two nematode populations

HG[a] Type indicator line number	Accession	Nematode Population	
		HG Type 0	HG Type 2.5.7
		—Female Index[b]—	
1	Peking	0	0
2	PI 88788	1	42
3	PI 90763	0	0
4	PI 437654	0	0
5	PI 209332	1	47
6	PI 89772	0	0
7	Cloud	3	62
	PI 507623	9	8
	PI 507613	3	14
	PI 342434	26	60
	PI 468916	2	18

[a]HG, *Heterodera glycines*

[b] Female index is percent of SCN cysts per root system, relative to the susceptible control (Lee 74) within the same experiment, for ten plants per genotype.

4.9 Supplemental Information

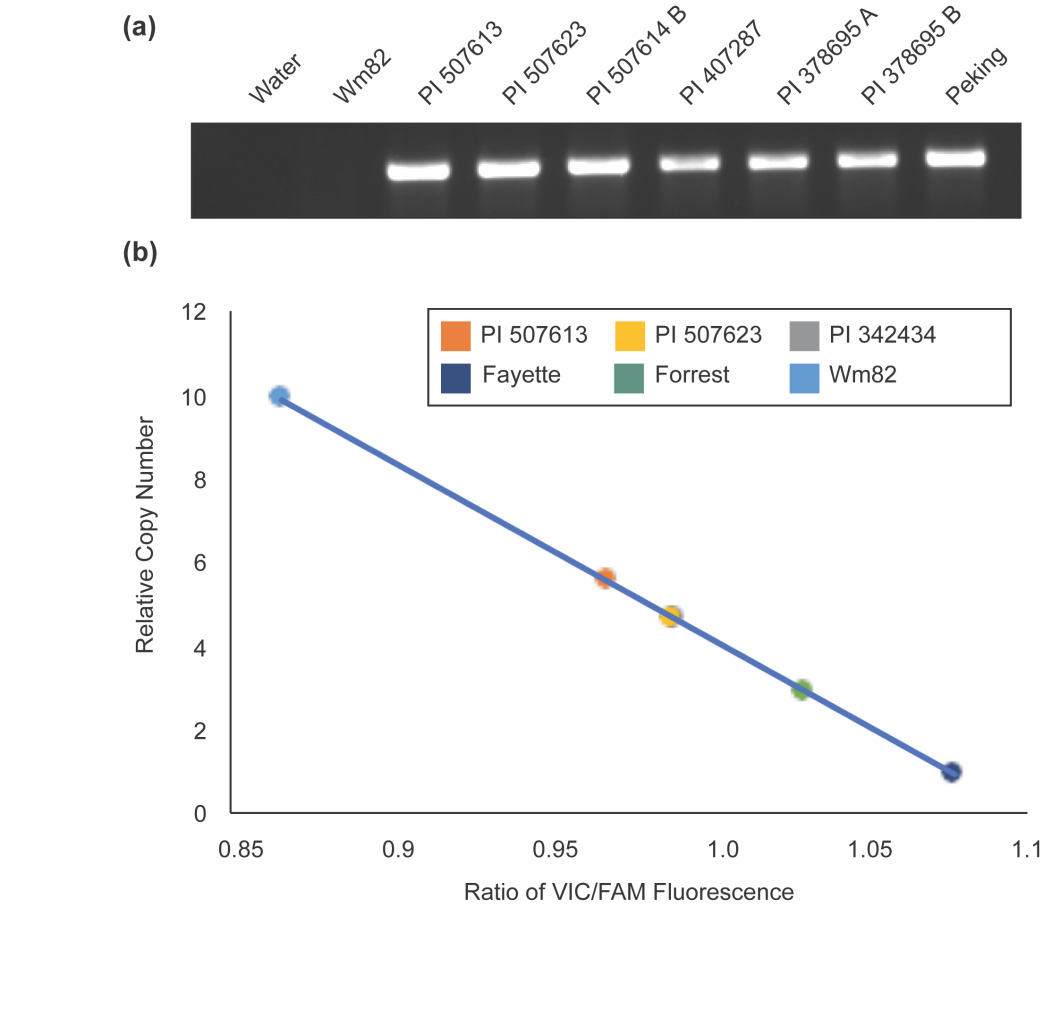
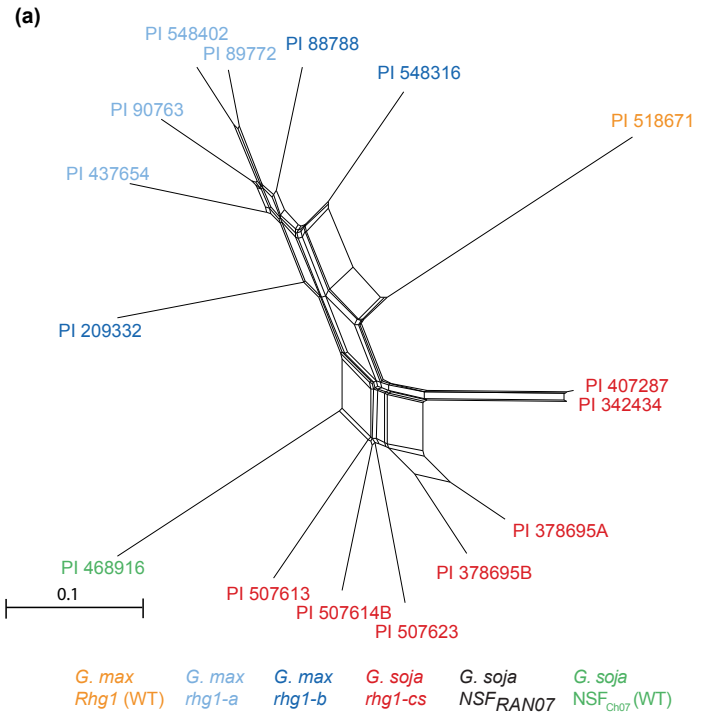


Figure S1. Validation of *Resistance to Heterodera glycines 1* (*Rhg1*) repeat junction and multi-copy *Rhg1* within *Glycine soja*. (a) Repeat junction was validated in a separate plant from those in Figure 1 using a different primer set. (b) Example of TaqMan assay on PI 507613, PI 507623, and PI 342434 along with *Glycine max* (Williams 82; Wm82), Fayette and Peking. Copy number estimated as the ratio of fluorescence of VIC (probe anneals *Rhg1* repeat) to FAM (probe anneals to single copy homeolog on chromosome 11). Blue line is the line of regression generated from Wm82, Peking and Fayette (blue line) ratios. Orange=PI 507613, Grey=PI

342434, Yellow=PI 507623, Fayette=dark blue, Forrest=green, Wm82=cyan. Note however that highly variable results were obtained between separate experiments.


(a)
PI 548402
PI 89772
PI 88788
PI 548316
PI 90763
PI 518671
PI 437654
PI 209332
PI 407287
PI 342434
PI 378695A
PI 468916
PI 378695B
0.1
PI 507613
PI 507614B
PI 507623
G. max Rhg1 (WT)
G. max rhg1-a
G. max rhg1-b
G. soja rhg1-cs
G. soja NSF RAN07
G. soja NSF Ch07 (WT)

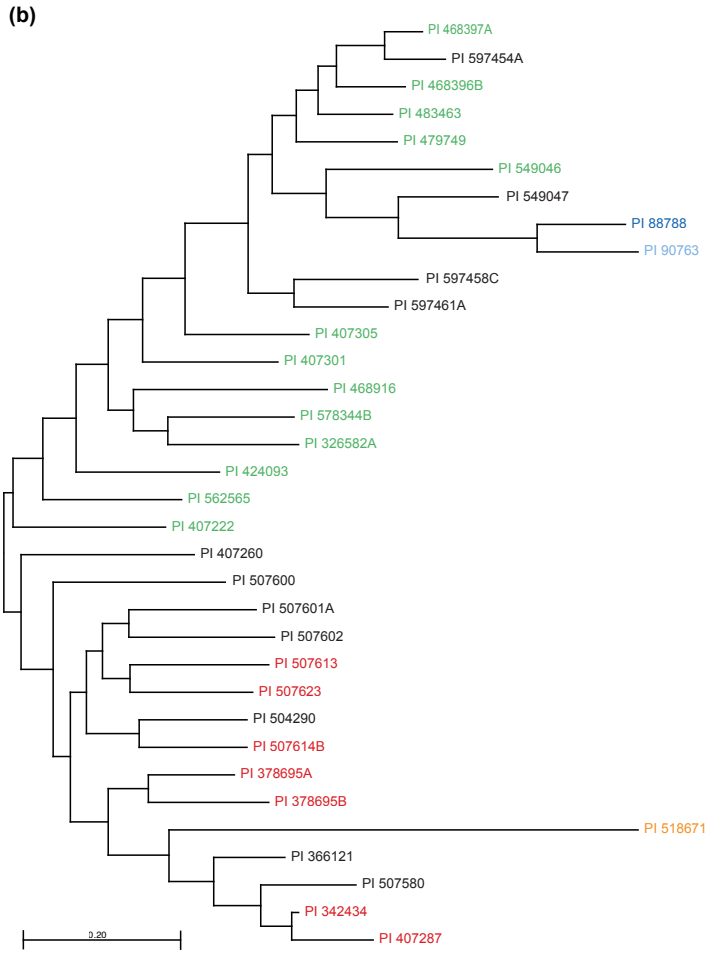

(b)
PI 468397A
PI 597454A
PI 468396B
PI 483463
PI 479749
PI 549046
PI 549047
PI 88788
PI 90763
PI 597458C
PI 597461A
PI 407305
PI 407301
PI 468916
PI 578344B
PI 326582A
PI 424093
PI 562565
PI 407222
PI 407260
PI 507600
PI 507601A
PI 507602
PI 507613
PI 507623
PI 504290
PI 507614B
PI 378695A
PI 378695B
PI 518671
PI 366121
PI 507580
PI 342434
PI 407287
0.20

Figure S2. *rhg1-cs*-containing G*lycine soja* form a sub-group within *G. soja* and *G. max*.

(a) Unrooted phylogeny generated from whole-genome single nucleotide polymorphism data comparing *rhg1-cs*-containing *G. soja* (red) to *G. max* soybean cyst nematode indicator lines from Table 1 (*rhg1-a* in light blue and *rhg1-b* in dark blue). (b) Phylogenetic tree generated from whole-genome SNP data (SoySNP50K) comparing *G. soja* that carry *rhg1-cs* and NSF_{RAN07} (red) to previously published *G. soja* of agronomic interest, that do not carry either *rhg1-cs* or NSF_{RAN07} (green). NSF_{RAN07}-containing *G. soja* that lack *rhg1-cs* marked with black, and Williams 82 (WT) in orange.

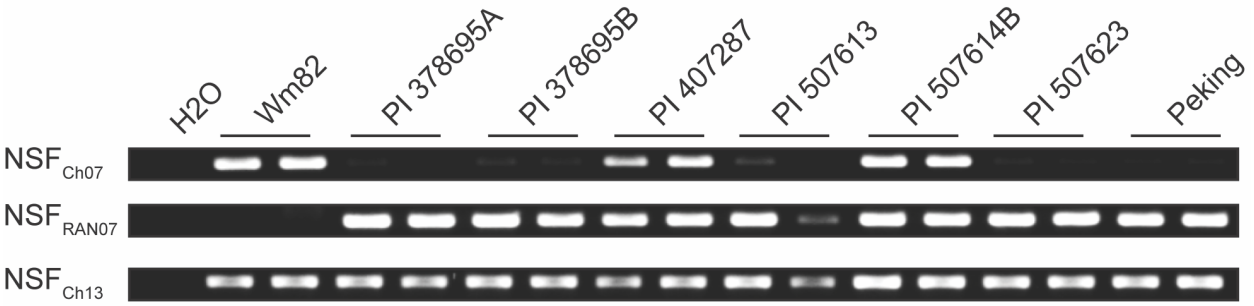


Figure S3. Multicopy *rhg1-cs Glycine soja* contain NSF_{RAN07}. Presence of the R_4Q polymorphism is verified in *G. soja* containing multicopy *Rhg1* (middle) relative to the wild-type R_4 (NSF_{Ch07}; top), or to chromosome 13 NSF (bottom). Data generated using allele-specific primers (Table S2). Pairs of bands are replicate PCR reactions for separate genomic DNA samples from the same accession.

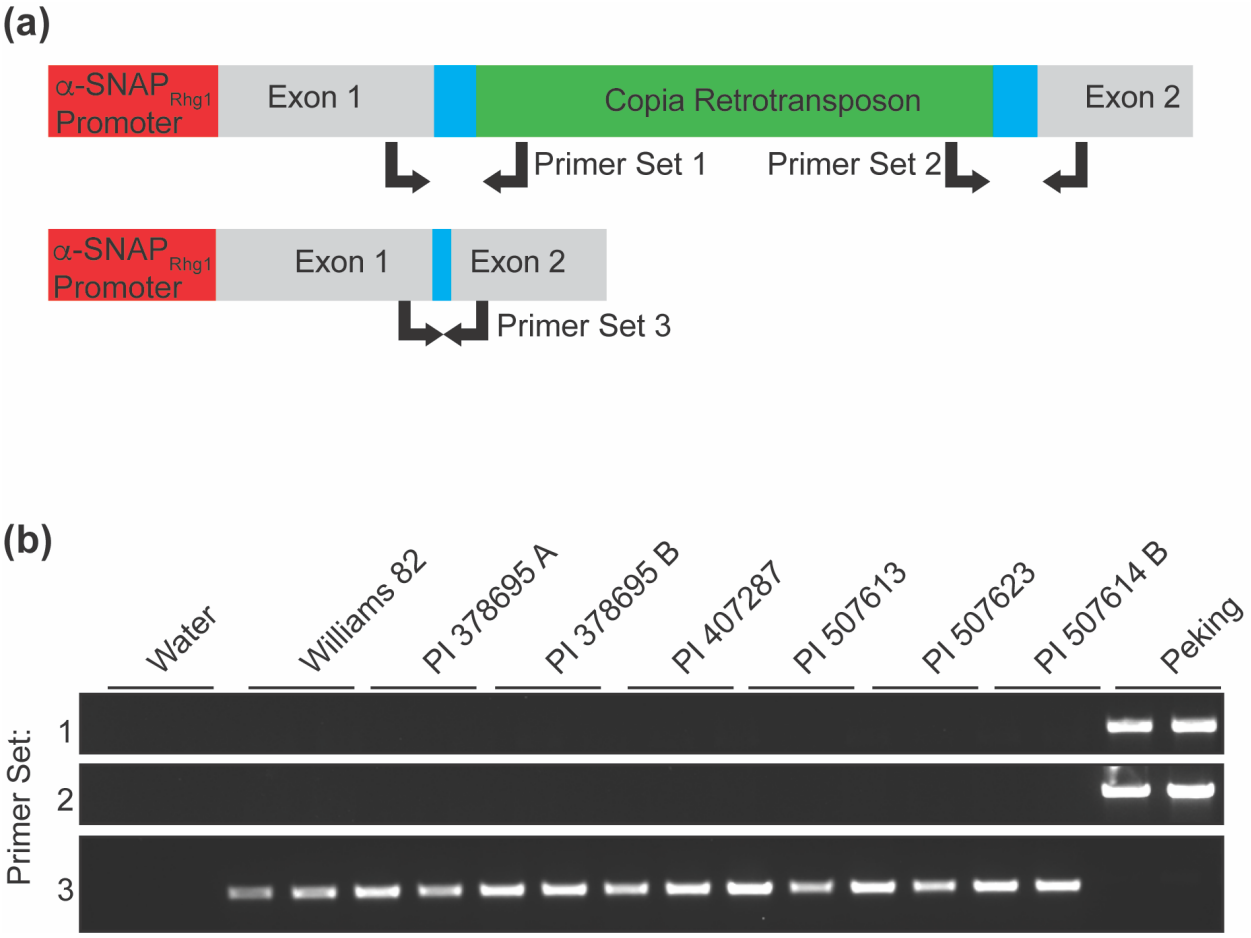


Figure S4. *rhg1*-cs-containing *Glycine soja* lack the RAC *rhg1-a*-associated retrotransposon. (a) Schematic of *Rhg1* α-SNAP copia (RAC) retrotransposon integration within known low-copy *Rhg1* soybean varieties. Primers used in (b) against the unique 5' and 3' junctions as well as across the entire first intron are also shown. (b) Gel showing amplification of the junctions as well as across the first intron in *G. max* cvs. Peking (a low-copy variety), Williams 82 (a single-copy variety) and the *rhg1*-cs *G. soja* accessions. Paired samples are replicate PCR reactions for different genomic DNA preparations of the same accession.

```
Wm82          AGATAGCTCGCCAATCCCTCAACAATAATTTGCTGAAGTATGGAGTTAAAGGACACCTGC  60
PI_507623     AGATAGCTCGCCAATCCCTCAACAATAATTTGCTGAAGTATGGAGTTAAAGGACACCTGC  60
PI_209332     AGATAGCTCGCCAATCCCTCAACAATAATTTGCTGAAGTATGGAGTTAAAGGACACCTGC  60
Peking        AGATAGCTCGCCAATCCCTCAACAATAATTTGCTGAAGTATGGAGTTAAAGGACACCTTC  60

Wm82          TTAATGCTGGCATCTGCCAACTCTGTAAAGGGGATGTTATTGCTGTAACCAATGCATTAG  120
PI_507623     TTAATGCTGGCATCTGCCAACTCTGTAAAGGGGATGTTATTGCTGTAACCAATGCATTAG  120
PI_209332     TTAATGCTGGCATCTGCCAACTCTGTAAAGGGGATGTTATTGCTGTAACCAATGCATTAG  120
Peking        TTAATGCTGGCATCTGCCAACTCTGTAAAGAGGAGGTTGTTGCTATAACCAATGCATTAG  120

Wm82          AACGATATCAGGTCTAAGTTTTTTCAATAGTCAACTTCTTGTGACTGGACAGCTTATTTG  180
PI_507623     AACGATATCAGGTCTAAGTTTTTTCAATAGTCAACTTCTTGTGACTGGACAGCTTATTTG  180
PI_209332     AACGATATCAGGTCTAAGTTTTTTCAATAGTCAACTTCTTGTGACTGGACAGCTTATTTG  180
Peking        AACGATATCAGGTCTAAGTTTTTTCAATAGTCAACTTCTTGTGACTGGACAGCTTATTTG  180

Wm82          TTGCTCAATTATTTAGATATGTTTTTATTTTGCAGGAACTGGATCCAACATTTTCGGGAA  240
PI_507623     TTGCTCAATTATTTAGATATGTTTTTATTTTGCAGGAACTGGATCCAACATTTTCAGGAA  240
PI_209332     TTGCTCAATTATTTAGATATGTTTTTATTTTGCAGGAACTGGATCCAACATTTTCGGGAA  240
Peking        TTGCTCAATTATTTAGATATGTTTTTATTTTGCAGGAACTGGATCCAACATTTTCAGGAA  240

Wm82          CACGTGAATATAGATTTTTGGCGGTAGGTCACTAGTTTTGGTATTTCGTTATTATTTTTT  300
PI_507623     CACGTGAATATAGATTTTTGGCGGTAGGTCACTAGTTTTGGTATTTCGTTATTATTTTTT  300
PI_209332     CACGTGAATATAGATTTTTGGCGGTAGGTCACTAGTTTTGGTATTTCGTTATTATTTTTT  300
Peking        CACGTGAATATAGATTTTTGGCGTTAGGTCACTAGTTTTGGTATTTCGTTATTATTTTTT  300

Wm82          TATTTCCAAGTAAATTGGATTAGAATATTTGAACTTCTTGGTTGCTGTCTCCTGGGTCAT  360
PI_507623     TATTTCCAAGTAAATTGGATTAGAATATTTGAACTTCTTTGTAGCTGTCTCCTGGGTCAT  360
PI_209332     TATTTCCAAGTAAATTGGATTAGAATATTTGAACTTCTTGGTTGCTGTCTCCTGGGTCAT  360
Peking        TATTTCCAAGTAAATTGGATTAGAATATTTGAACTTCTTTGTAGCTGTCTCCTGGGTCAT  360

Wm82          AATGTTTTATTATATTTTGGTATTAGCATAGCATTGTGATAGCACTATTACTATTTTGTT  420
PI_507623     AATGTTTTATTGTATTTTGGTATTAGCATAGCATTGTGATAGCACTATTACTATTTTGTT  420
PI_209332     AATGTTTTATTATATTTTGGTATTAGCATAGCATTGTGATAGCACTATTACTATTTTGTT  420
Peking        AATGTTTTATTGTATTTTGGTATTAGCATAGCATTGTGATAGCACTATTGCTATTTTGTT  420

Wm82          TGCTGATTCACTAGTACATTTGCCAGATGAAACTGACATTTTTTTTAATCCTGGTGGATA  480
PI_507623     TGCTGATTCACTAGTACATTTGCCAGATGAAACTGACATTTTTTTTAATCCTGGTGGATA  480
PI_209332     TGCTGATTCACTAGTACATTTGCCAGATGAAACTGACATTTTTTTTAATCCTGGTGGATA  480
Peking        TGCTGATTCACTAGTACATTTGCCAGATGAAACTGACATTTTTTTTAATCCTGGTGGATA  480

Wm82          GGACATTGCTGCTGCAATTGATGAAGAAGATGTTGCGAAGTTTACTGATGTTGTCAAGGA  540
PI_507623     GGACATTGCTGCTGCAATTGATGAAGAAGATGTTGCGAAGTTTACTGATGTTGTCAAGGA  540
PI_209332     GGACATTGCTGCTGCAATTGATGAAGAAGATGTTGCGAAGTTTACTGATGTTGTCAAGGA  540
Peking        GGACATTGCTGCTGCAATTGATGAAGAAGATGTTGCGAAGTTTACTGATGTTGTCAAGGA  540

Wm82          ATTTGATAGCATGACCCCTCTGGTAAGCTCCAAAAGTTGTTAAGTAGGATAACTTCTAGT  600
PI_507623     ATTTGATAGCATGACCCCTCTGGTAAGCTCCAAAAGTTGTTAAGTAGGATAACTTCTAGT  600
PI_209332     ATTTGATAGCATGACCCCTCTGGTAAGCTCCAAAAGTTGTTAAGTAGGATAACTTCTAGT  600
Peking        ATTTGATAGCATGACCCCTCTGGTAAGCTCCAAAAGTTGTTAAGTAGGATAACTTCTAGT  600

Wm82          GGTATTTAACAAAAATACTTCCACTGTATTTTTTATCCACATTTTATACAA  651
PI_507623     GGTATTTAACAAAAATACTTCCACTGTATTTTTTATCCACATTTTATACAA  651
PI_209332     GGTATTTAACAAAAATACTTCCACTGTATTTTTTATCCACATTTTATACAA  651
Peking        GGTATTTAACAAAAATACTTCCACTGTATTTTTTATCCACATTTTATACAA  651
```

Figure S5. Multicopy *rhg1*-cs-containing *Glycine soja* do not contain the chromosome 11 α-SNAP$_{Ch11}$ intron retention allele. Genomic sequence comparison of the α-SNAP$_{Ch11}$ region surrounding the previously described intron retention allele. Note that *G. max* PI 209332 carries *rhg1-b*. High copy (HC)-like single nucleotide polymorphisms highlighted in red and Peking-like SNPs highlighted in blue. The black box indicates the intron retention-associated SNP. Note that the sequence for *G. soja* PI 507623 was identical to the sequences for PCR-amplified genomic DNA obtained for the other six *rhg1-cs* accessions PI 342434, PI 378695A, PI 378695B, PI 407287, PI 507613 and PI 507614B. The SoySNP50K ss715610416 site associated with this intron-retention allele (Table S1; Bayless et al., 2018) is at a closely linked site not examined in the present experiment.

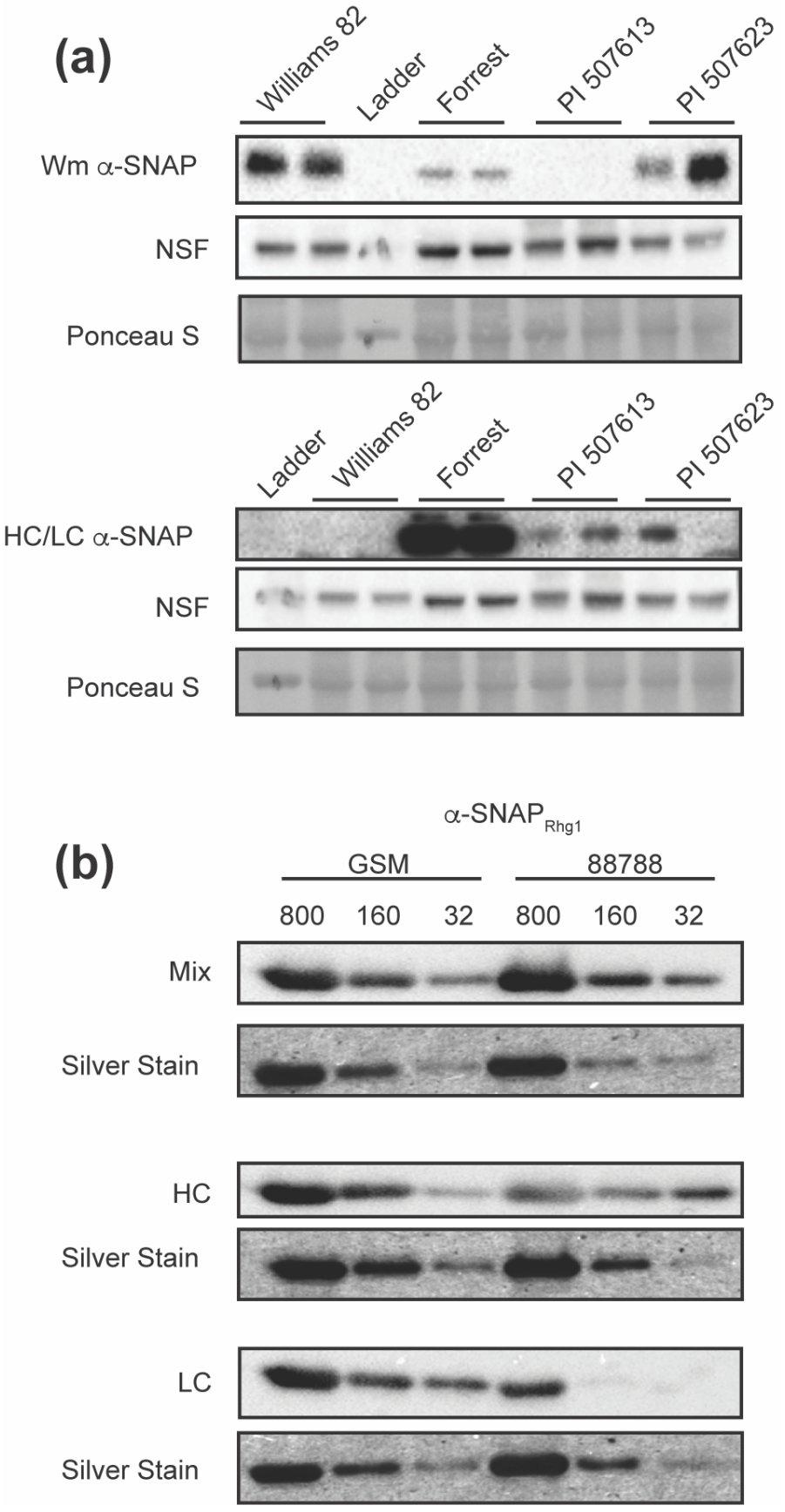

(a)
Williams 82
Ladder
Forrest
PI 507613
PI 507623
Wm α-SNAP
NSF
Ponceau S
Ladder
Williams 82
Forrest
PI 507613
PI 507623
HC/LC α-SNAP
NSF
Ponceau S
(b)
α-SNAP Rhg1
GSM
88788
800
160
32
800
160
32
Mix
Silver Stain
HC
Silver Stain
LC
Silver Stain

Figure S6. Some *rhg1-cs*-containing G*lycine soja* have lower apparent expression of multicopy-encoded α-SNAP$_{Rhg1}$Gsm and/or wild-type α-SNAP in developing trifoliates.

(a) Immunoblot of protein extracts from developing trifoliates, showing NSF, α-SNAP$_{WT}$ and α-SNAP$_{Rhg1}$ abundance in select *G. soja* varieties along with Williams 82 (Wm82) and Forrest controls. Polyclonal antibody used is listed to left of each set of lanes. Ponceau S shown as overall protein loading control. (b) Immunoblots showing relative detection of purified α-SNAP$_{Rhg1}$Gsm by custom antibodies, in comparison to detection of high copy α-SNAP (α-SNAP$_{Rhg1}$HC) by the same antibodies. Polyclonal antibody used is listed to left of each set of lanes; 'Mix' is anti-HC + anti-LC. Silver stain shown as loading control. The numbers 800, 160, 32 indicate pg of purified recombinant α-SNAP protein loaded in that lane.

Table S1: Single nucleotide polymorphism identity of select *Glycine max* and *Glycine soja* accessions[1]			
Accession	ss715597431 (NSF_{RAN07}) SNP identity	ss715610416 (α-$SNAP_{Ch11}$-IR) SNP identity[2]	ss715606985 (RAC) SNP identity
PI 342434	T	C	G
PI 378695B	U	U	G
PI 366121	T	U	G
PI 378695A	T	U	G
PI 407260	T	U	G
PI 407287	T	U	G
PI 504290	T	U	G
PI 507580	T	U	G
PI 507600	T	U	G
PI 507601A	T	U	G
PI 507602	T	U	G
PI 507603	T	U	G
PI 507604	T	U	G
PI 507608	T	U	G
PI 507613	T	U	G
PI 507614B	T	C	G
PI 507623	T	U	G
PI 549047	T	U	G
PI 597454A	T	U	G
PI 597458C	T	U	G
PI 597461A	T	C	G
PI 597461C	T	U	G
PI 518671 (Wm82)	C	C	G
PI 88788	T	C	G
PI 548402 (Peking)	T	T	A

[1]*Glycine soja* accessions with single nucleotide polymorphism indicative of R_4Q NSF_{RAN07} mutation are shown (21 out of 1169 *Glycine soja* accessions screened using SoySNP50K data), along with PI 342434 and *G. max* controls on last three lines. Non-functional intron-retention mutation (a-$SNAP_{Ch11}$-IR), and *Rhg1* α-SNAP Copia (RAC) SoySNP50K SNP genotypes shown in final two columns. U indicates uncalled SNP.

[2]See also Figure S5 for DNA sequence data obtained for same region.

Chapter 5: Conclusions and Future Direction

The work we present here elucidates many facets of the soybean cyst nematode (SCN)-soybean pathosystem as well as contributes to our understanding on the evolution of a major SCN resistance QTL, *Rhg1*. In Chapter 2, we discover a novel NSF chaperone, termed NSF_{RAN07}, that both co-segregates with *Rhg1* as well as is necessary for viability of resistance-type *Rhg1*-containing *Glycine max* (Bayless et al., 2018). The work presented in this chapter develops novel SNP markers that may be used to screen soybean germplasm for both NSF_{RAN07}, as well as a *rhg1-a*-associated α-$SNAP_{Ch11}$ intron retention allele. This work answers a question previously posed for *rhg1-a* haplotype soybeans, namely: in the absence of Wm82-type α-SNAP, how are these soybeans viable? Through mutation in the N-domain (the NSF/α-SNAP interacting surface), NSF_{RAN07} is able to accommodate both *rhg1-a* and *rhg1-b*-type α-SNAPs. The conservation of NSF_{RAN07} between *rhg1-a* and *rhg1-b* further suggests a shared origin for those *Rhg1* haplotypes, one that may have arisen in *G. soja*, the wild ancestor of *G. max* (see Chapter 4).

Future experiments could explore the diversity of novel NSFs within exotic soybean germplasm. In Chapter 4, we utilized the NSF R_4Q SNP to choose lines to screen for *Rhg1* within *G. soja*. However, from the above work, this polymorphism is not necessary for viability of *G. max* in recombinant inbred line (RIL) populations; rather, $\hat{}_{116}F$ and $M_{181}I$ were found to be consistently associated with *Rhg1* in these populations. Therefore, additional work may be done to screen first for *Rhg1* (either guided or unguided by previously characterized SNPs) followed by characterization of the NSF; that is, looking for absence of amplification by the R_4Q-

specific primer set used in Chapters 2 and 4. Such experiments will help to elucidate the evolution of *Rhg1*, i.e., whether *Rhg1* necessitated the evolution of NSF_{RAN07} or whether the preemergence of NSF_{RAN07} provided the conditions for *Rhg1* evolution. That is, are there other NSFs in annual *Glycine* populations that can both allow for presence of *Rhg1*? Furthermore, such germplasm may be a valuable source of additional NSF alleles should NSF_{RAN07} be a target of SCN virulence.

In Chapter 2 we also discuss the presence of a *rhg1-a*-associated SNP that corresponds to an α-$SNAP_{Ch11}$ intron retention allele, with consequently reduced Wm82 α-SNAP levels. This is significant as there is evidence to suggest that *rhg1-a*-type α-SNAP interacts both with $SHMT_{Rhg4}$ (the serine hyrdoxymethyltransferase found in *rhg1-a G. max* accessions) and a pathogen resistance protein to promote resistance to HG type 0 and HG type 2.5.7 nematode populations (Lakhssassi et al., 2020). The above model of resistance is supported by the observation that all three proteins are necessary to induce cell death and for complete physical interaction. The mechanism underlying this resistance is unknown, though one possibility suggests that sequestration of $SHMT_{Rhg4}$ by the complex might result in reduction in one-carbon metabolism, with subsequent reduction in nucleotide and amino acid anabolism and/or methylation. Moreover, the interaction of $SHMt_{Rhg4}$ with α-SNAP is reminiscent of glyceraldehyde 3-phosphate dehydrogenase (GAPDH), which interacts with tethering components during early vesicle trafficking steps (Tisdale et al., 2004). Is it possible that the $SHMT_{Rhg4}$ modulates vesicle trafficking through its activity with α-SNAP?

Future work may here be directed toward elucidating the interaction between *rhg1-b* and/or Wm82-type *Rhg1* α-SNAP with $SHMT_{Rhg4}$, whereby transgenic plants may be made that

contain a Wm82-type α-SNAP and/or $SHMT_{Rhg4}$. Resistance assays on the resulting plants will determine whether Wm82-type α-SNAP abrogates the interaction between *rhg1-a*-type α-SNAP and $SHMT_{Rhg4}$, with concomitant reduction in HG type 2.5.7 resistance. To address the sequestration hypothesis, availability of $SHMT_{Rhg4}$ active sites can be probed with fluorescent probes (Nanoka et al., 2019) in different *Rhg1* α-SNAP backgrounds or through the course of infection. To determine whether the $SHMT_{Rhg4}$ modulates vesicle trafficking, directly or indirectly, sec-GFP experiments can be performed wherein differences in secretion may be observed by GFP signal retention in different $SHMT_{Rhg4}$ (resistant or susceptible type) backgrounds during infection.

The role of NSFs in nematode infection could be explored. There is evidence to suggest that NSF protein abundance increases in syncytia of *rhg1-b* soybeans; however, whether this is NSF_{Ch13} or NSF_{RAN07} could not be resolved using our custom antibody (Bayless et al., 2016). One methodology which might be employed is generating transgenic plants conditionally expressing or overexpressing either NSF_{RAN07}, NSF_{Ch13} or NSF_{Ch07} in *rhg1-a* or *rhg1-b* soybean backgrounds. Subsequent resistance assays can then be used to determine the relative importance of these NSFs during nematode infection. More nuanced interaction experiments may be performed through utilizing proximity labeling techniques during SCN infection (Mair et al., 2019; Zhang et al., 2019). Specific interaction between SCN effector molecules and partner α-SNAPs and NSFs is discussed in Appendix 1.

That NSF_{RAN07} better interacts with α-$SNAP_{Rhg1}$ from *rhg1-a*, *rhg1-b*, and *rhg1-cs* varieties than any interacts with NSF_{Ch07} in vitro is complicated by the observation that NSF_{Ch07} is only moderately less cytoprotective than NSF_{Ran07} in *N. benthamiana*. This suggests that the ATPase

activity of NSF_{Ch07} is comparable to if not greater than NSF_{RAN07}. Such a hypothesis may be tested on the α-$SNAP_{Rhg1}$ from Williams 82, which is able to bind both NSF_{Ch07} and NSF_{RAN07} and ATPase activity tested in vitro using established malachite green assays for example, as described in Carter and Karl (1982) and Rancour *et al.* (2004).

In Chapter 3, we characterize a ty-copia-type retrotransposon, termed RAC, found in *rhg1-a* soybean varieties (Bayless et al., 2019). In this work, we present the discovery of a fully-intact and transcriptionally active ty-copia type retrotransposon located antisense in the first intron of α-SNAP in *rhg1-a* soybean accessions. Additionally, we develop a novel SNP marker for this retroelement that associates both with previously characterized α-$SNAP_{Ch11}$ intron retention allele and *rhg1-a* SNP markers in highly resistant soybean lines. We further show that this RAC element is transcriptionally active but does not seem to impact α-SNAP protein abundance or alternative splicing of the *GmSNAP18* transcript in the absence of SCN. Finally, we show that like *rhg1-b*, *rhg1-a* α-SNAP abundance increases in syncytia, suggesting a similar albeit non-identical mechanism of conferring SCN resistance.

Future work may be directed towards elucidating resistance mechanisms underlying putatively *rhg1-a* but non-RAC-containing soybean accessions. From the work presented here, there are many such soybean accessions in the moderately resistant category. Whether such accessions are truly lacking RAC and/or do not have a signature of prior RAC integration, are *rhg1-a*, and are resistant should first be elucidated using PCR and nematode resistance assay methods. If these accessions meet these criteria, then future work may be directed towards understanding additional loci that might be contributing to resistance, e.g., through conventional breeding and mapping or RNA-seq based methods. This is especially prudent as

there is at least one putatively *rhg1-a*-type as well as a number of *Rhg1*-containg *G. soja* that are both highly resistant to HG type 2.5.7 and which lack RAC (see Chapter 4).

Retrotransposons are known to be responsive to both biotic and abiotic stresses. Retroelements in promoter or intronic regions of pathogen-responsive genes have been shown to be necessary for their function (Hayashi et al., 2009; Tsuchiya and Eulgem, 2013). It remains to be explored whether the same holds true for the transposable element described here. Under SCN infection conditions, does the retroelement change expression itself, or alter expression of other genes by acting as an enhancer to any of the genes in the *Rhg1* locus? As a first step, qPCR or western blots may be utilized to determine whether any of the *rhg1-a* (specifically *GmSNAP18*) genes change expression in response to SCN, or when expressed heterologously in *rhg1-b* or Wm82 soybeans using native constructs containing or not containing the RAC. Does the position of the retroelement play a role in modulating expression of proximal genes? The RAC is a proximal element to the promoter of *Glyma.18G022400* (amino acid permease). Expression of the amino acid permease could be observed by qPCR in Williams 82 roots that have an artificial construct with and without RAC integration upstream of the gene at its native locus. Intronic and promoter-localized transposons have been shown to modulate gene expression. Would the RAC differentially modify expression of *GmSNAP18* if it were introduced as a promoter element rather than an intronic retroelement? Such questions may be answered by generating Williams 82 transgenic roots where the RAC element incorporated at various proximities to the transcription start site of *Glyma.18G022500* from *rhg1-a* accessions and α-SNAP$_{Rhg1}$LC measured by immunoblots or qPCR.

In Forrest soybeans, the retroelement was shown to be constitutively active, presumably through transcription from the LTRs. Broadly, what is the regulation of RAC, and can this regulation be exploited for expression of executor genes, or genes of other agronomic interest? What are the constituent responsive elements within the LTR of the RAC element? One way to answer this is through "promoter bashing" experiments, wherein various sub-portions of the LTR are cloned upstream of a reporter gene (e.g., GFP). The promoter-GFP construct may then be used to generate transgenic roots, and GFP expression measured under various abiotic or biotic stressors (e.g., SCN). At baseline, how does transcription from the LTR compare to transcription from other strong or conditional promoters? It is an intriguing possibility to utilize LTRs as promoters to drive genes of interest. The most notable use of this being the glucocorticoid responsive element of MMTV-LTR, and its use in expressing the human epithelial cell adhesion molecule (Otten et al., 1988; Vercollone et al., 2015). LTRs may expand the library of available promoters in genetic engineering experiments, albeit while potentially encouraging recombination.

In Chapter 4, we discovered a novel multi-copy *Rhg1* haplotype, termed *rhg1-cs*, in the wild relative of soybean, *G. soja*. The *rhg1-cs*-containing *G. soja* have a novel allele of *GmSNAP18*, whose protein product behaves similarly as the α-SNAPs in *rhg1-a* and *rhg1-b G. max*. Characterization of a couple of these *G. soja* accessions suggested that *rhg1-cs* arose prior to the divergence between *rhg1-a* and *rhg1-b*. This work also begins to answer the question as to origin of the RAC element: namely did RAC arise prior to the split of *rhg1-a* and *rhg1-b* or emerge in *rhg1-a*? From our study of *Rhg1*-containing *G. soja*, an emerging hypothesis is that the RAC element did not arise in *G. soja*, rather that it arose specifically in *rhg1-a*-containing

soybean. In addition to answering evolutionary questions, the work presented here discovered new germplasm that is highly resistant to HG type 2.5.7 nematode populations, but which does not seem to utilize the known sources of this resistance in *G. max*.

As the *G. soja* screened here do not seem to rely on known sources of HG type 2.5.7 resistance, an obvious question is: what is the mechanism underlying resistance to this nematode population in *G. soja*? It does not seem to be the case that resistance is due to nonhost resistance in *G. soja*; rather, there is likely a complex transcriptional response that underlies resistance (Khan et al., 2004; Klink et al., 2005; Klink et al., 2007a; Klink et al., 2007b; Klink et al., 2009; Klink et al., 2010; Kandoth et al., 2011; Zhang et al., 2017). Future work could be directed towards understanding resistance in *G. soja* though conventional breeding by looking for segregation of resistance in the LD10-10198 X PI 507623 F1 or F2 populations generated for this chapter. Particularly, rather than looking for segregation of *Rhg1* in the F2 population and testing these for resistance, all F2 seeds could be tested for resistance, followed by mapping. Additionally, future work may be directed to uncovering potential transcriptional networks that underlies HG type 2.5.7 resistance in the resistant *G. soja*, for example through RNAseq, though post-transcriptional regulation of vesicle trafficking may play a role in resistance.

Indeed, it has been an observation that the α-SNAP$_{Rhg1}$WT migrates with greater mobility than the α-SNAP$_{Rhg1}$ from *rhg1-a*, *rhg1-b* or *rhg1-cs* containing accessions. The single amino acid difference between the α-SNAPs is unlikely to account for this mobility shift. Rather, it could be possible that there is processing of the α-SNAP$_{WT}$ in bacteria that could account for reduced interaction with NSF$_{Ch07}$. Efforts should be made to analyze the peptide from this a-

$SNAP_{WT}$ to determine if there is processing, and there should be efforts made to determine whether the same observation is made in soybeans. Here, transgenic *rhg1-a* soybeans should be generated which also contain an α-$SNAP_{WT}$, and combined WT/LC immunoblots run to compare mobility of α-SNAP these roots that in *rhg1-a* (no α-$SNAP_{WT}$) and Williams 82.

Above, we describe *Rhg1* within annual *Glycine* species. The conservation of *Rhg1* between *G. soja* and *G. max* suggests that the origin of *Rhg1*-based SCN resistance may have arisen even earlier, within the perennial *Glycine* species. Indeed, perennial *Glycine* species, more than annual relatives, tend to have broader resistance, including resistance to the highly virulent HG type 1.2.3.4.5.6.7 nematode population (Riggs et al., 1998; Bauer et al., 2007; Wen et al., 2017; Herman et al., 2020). To date, in the most comprehensive study on perennial *Glycine* species, Bauer et al. (2007) screened 491 accessions from 12 perennial *Glycine* species. From this work 10 of the species had resistance to HG type 0 nematode populations. Do perennial *Glycine* species have *Rhg1*? As perennial *Glycine* species are not in the SoySNP50K dataset, a subset of resistant accessions can be screened for presence of multi-copy *Rhg1* and other co-segregating loci. Presence of multi-copy *Rhg1* would indicate an even earlier origin of *Rhg1* and *Rhg1*-based resistance within the *Glycine* genus. A screen of perennial *Glycine* species is described in appendix 5.

The discovery of a novel α-SNAP in *G. soja* raises the question: within *Rhg1*, does a particular α-SNAP repertoire influence the resistance outcome? There is a growing body of evidence that α-SNAP and their interactions with SNAREs or SNARE-like proteins form a core component of SCN-*G. max* interactions (Bekal et al., 2015; Gardner et al., 2018; Dong et al., 2020). As noted above, there is even suggestion of $SHMT_{Rhg4}$ interacting with *rhg1-a*-type α-

SNAP, further implicating specificity of α-SNAP as an important component in the SCN-*G. max* pathosystem (Lakhssassi et al., 2020). Introducing novel α-SNAPs with mutations along putative interacting surfaces with NSF, SNAREs, and/or $SHMT_{Rhg4}$ will go a long way to elucidating the mechanisms of *rhg1-a* and *rhg1-b* resistance, as well as concomitantly producing α-SNAPs which may confer enhanced or novel resistance. Both creating the point mutations at the α-SNAP/NSF interaction surface and preliminary characterization of these mutants are described in Appendix 3.

What other utility does *Rhg1* provide? The ability of *Rhg1* to confer resistance to cyst nematodes in diverse plant species has been explored in common bean (*Phaseolus vulgaris*), *Arabidopsis thaliana* and potato [*Solanum tuberosum*; (Butler et al., 2019; Wen et al., 2019)]. Within common bean, genome wide association studies mapped a significant SNP associated with HG type 2.5.7 resistance near a region with genes orthologous to those found in *Rhg1* (Wen et al., 2019). However, sequencing of the α-SNAP-encoding gene at this locus did not reveal any differences between resistant and susceptible varieties, suggesting that HG type 2.5.7 resistance functions in a manner different than what is hypothesized to occur in *G. max* (appendix 5). Heterologous expression of *Rhg1* genes in both *A. thaliana* and potato improved resistance to each of their cognate *Heterodera* or *Globodera* cyst nematode pathogens (Butler et al., 2019). However, in these cases, there were significant agronomic consequences including reduced root and tuber masses (Butler et al., 2019). This is likely due to not including the NSF_{RAN07} as part of the transformation, and future work may be directed to exploring whether its inclusion can abrogate the negative effects.

Can *Rhg1* be used to confer resistance to non-nematode pathogens? Vesicle trafficking plays a central role in plant-pathogen interactions (Hoefle et al., 2008; Frey and Robatzek, 2009; Cai et al., 2019; Ekanyake et al., 2019). As cell death from disruption of vesicle trafficking takes 2-3 days, It is feasible that disrupting vesicle trafficking through activation of the *Rhg1*-type α-SNAP can confer specific or broad resistance to a variety of slow-growing pathogens. Enhancer trapping has been utilized to find plant genes that are specifically expressed in the presence of pathogens (Fridborg et al., 2004; Schroeder et al., 2016). Early work in enhancer trapping discovered *TMV Induced Gene* (*TRI*; *At1g78410*), which was upregulated in the presence of TMV, *Pseudomonas syringae* and *Hyaloperonospora arabidopsidis* (HPA; Fridborg et al., 2004). More recent work showed a specific upregulation of *Myosin Binding Protein 1* (*MyoB1*; *At1g08800*) during compatible *A. thaliana*-HPA interactions (Schroeder et al., 2016). Interestingly, MyoB1 is a protein involved in type XI myosin binding, proteins that play a role in conferring disease resistance through a unique vesicle trafficking pathway (Peremyslov et al., 2013; Yang et al., 2014). Design and construction of *proMyoB1*::α-SNAP is discussed in Appendix 2, and follow-up work to test the hypothesis that may be productive.

Description of novel loci co-segregating with *Rhg1*, discovery of novel and highly resistant germplasm within *G. soja* and finding additional application of *Rhg1* to other plant pathosystems will go a long way in providing growers, breeders and molecular biologists with exciting opportunities in the future to find ways to combat extensive crop loss caused by SCN and other plant pathogens.

5.2 References

Bauer, S., Hymowitz, T. and Noel, G. (2007). Soybean Cyst Nematode Resistance Derived from Glycine tomentella Amphiploid (G. max X G. tomentella) Hybrid Lines. *Nematropica*, 37(2), pp.277–285.

Bayless, A.M., Smith, J.M., Song, J., McMinn, P.H., Teillet, A., August, B.K. and Bent, A.F. (2016). Disease resistance through impairment of α-SNAP–NSF interaction and vesicular trafficking by soybean Rhg1. *Proceedings of the National Academy of Sciences*, 113(47), pp.E7375–E7382.

Bayless, A.M., Zapotocny, R.W., Grunwald, D.J., Amundson, K.K., Diers, B.W. and Bent, A.F. (2018). An atypical N-ethylmaleimide sensitive factor enables the viability of nematode-resistant Rhg1 soybeans. *Proceedings of the National Academy of Sciences*, 115(19), pp.E4512–E4521.

Bayless, A.M., Zapotocny, R.W., Han, S., Grunwald, D.J., Amundson, K.K. and Bent, A.F. (2019). The rhg1-a (Rhg1 low-copy) nematode resistance source harbors a copia-family retrotransposon within the Rhg1- encoded α-SNAP gene. *Plant Direct*, 3(8), p.e00164.

Bekal, S., Domier, L.L., Gonfa, B., Lakhssassi, N., Meksem, K. and Lambert, K.N. (2015). A SNARE-Like Protein and Biotin Are Implicated in Soybean Cyst Nematode Virulence. *PLOS ONE*, 10(12), p.e0145601.

Butler, K.J., Chen, S., Smith, J.M., Wang, X. and Bent, A.F. (2019). Soybean Resistance Locus Rhg1 Confers Resistance to Multiple Cyst Nematodes in Diverse Plant Species. *Phytopathology™*, 109(12), pp.2107–2115.

Cai, Q., He, B. and Jin, H. (2019). A safe ride in extracellular vesicles – small RNA trafficking between plant hosts and pathogens. *Current Opinion in Plant Biology*, [online] 52, pp.140–148. Available at: https://www.sciencedirect.com/science/article/pii/S1369526618301560?via%3Dihub [Accessed 21 Feb. 2020].

Carter, S.G. and Karl, D.W. (1982). Inorganic phosphate assay with malachite green: An improvement and evaluation. *Journal of Biochemical and Biophysical Methods*, 7(1), pp.7–13.

Dong, J., Zielinski, R.E. and Hudson, M.E. (2020). t-SNAREs Bind the Rhg1 α-SNAP and Mediate Soybean Cyst Nematode Resistance. *The Plant Journal*.

Ekanayake, G., LaMontagne, E.D. and Heese, A. (2019). Never Walk Alone: Clathrin-Coated Vesicle (CCV) Components in Plant Immunity. *Annual Review of Phytopathology*, 57(1), pp.387–409.

Frey, N.F. dit and Robatzek, S. (2009). Trafficking vesicles: pro or contra pathogens? *Current Opinion in Plant Biology*, 12(4), pp.437–443.

Fridborg, I., Williams, A., Yang, A., MacFarlane, S., Coutts, K. and Angell, S. (2004). Enhancer Trapping IdentifiesTRI, anArabidopsisGene Up-Regulated by Pathogen Infection. *Molecular Plant-Microbe Interactions*, 17(10), pp.1086–1094.

Gardner, M., Dhroso, A., Johnson, N., Davis, E.L., Baum, T.J., Korkin, D. and Mitchum, M.G. (2018). Novel global effector mining from the transcriptome of early life stages of the soybean cyst nematode Heterodera glycines. *Scientific Reports*, 8(1).

Hayashi, K. and Yoshida, H. (2009). Refunctionalization of the ancient rice blast disease resistance genePitby the recruitment of a retrotransposon as a promoter. *The Plant Journal*, 57(3), pp.413–425.

Herman, T.K., Han, J., Singh, R.J., Domier, L.L. and Hartman, G.L. (2020). Evaluation of wild perennial Glycine species for resistance to soybean cyst nematode and soybean rust. *Plant Breeding*.

Hoefle, C. and Hückelhoven, R. (2008). Enemy at the gates: traffic at the plant cell pathogen interface. *Cellular Microbiology*, 10(12), pp.2400–2407.

Kandoth, P.K., Ithal, N., Recknor, J., Maier, T., Nettleton, D., Baum, T.J. and Mitchum, M.G. (2011). The Soybean Rhg1 Locus for Resistance to the Soybean Cyst Nematode Heterodera glycines Regulates the Expression of a Large Number of Stress- and Defense-Related Genes in Degenerating Feeding Cells. *Plant Physiology*, 155(4), pp.1960–1975.

Khan, R., Alkharouf, N., MacDonald, M., Chouikha, I., Meyer, S., Grefenstette, J., Knap, H. and Matthews, B. (2004). Microarray Analysis of Gene Expression in Soybean Roots Susceptible to the Soybean Cyst Nematode Two Days Post Invasion. *The Journal of Nematology*, 36(3), pp.214–248.

Klink, V.P., Alkharouf, N., MacDonald, M. and Matthews, B. (2005). Laser Capture Microdissection (LCM) and Expression Analyses of Glycine max (Soybean) Syncytium Containing Root Regions Formed by the Plant Pathogen Heterodera glycines (Soybean Cyst Nematode). *Plant Molecular Biology*, 59(6), pp.965–979.

Klink, V.P., Hosseini, P., Matsye, P., Alkharouf, N.W. and Matthews, B.F. (2009). A gene expression analysis of syncytia laser microdissected from the roots of the Glycine max (soybean) genotype PI 548402 (Peking) undergoing a resistant reaction after infection by Heterodera glycines (soybean cyst nematode). *Plant Molecular Biology*, 71(6), pp.525–567.

Klink, V.P., Hosseini, P., Matsye, P.D., Alkharouf, N.W. and Matthews, B.F. (2010). Syncytium gene expression in Glycine max[PI 88788] roots undergoing a resistant reaction to the parasitic nematode Heterodera glycines. *Plant Physiology and Biochemistry*, 48(2–3), pp.176–193.

Klink, V.P., Overall, C.C., Alkharouf, N.W., MacDonald, M.H. and Matthews, B.F. (2007a). A time-course comparative microarray analysis of an incompatible and compatible response by Glycine max (soybean) to Heterodera glycines (soybean cyst nematode) infection. *Planta*, 226(6), pp.1423–1447.

Klink, V.P., Overall, C.C., Alkharouf, N.W., MacDonald, M.H. and Matthews, B.F. (2007b). 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(6), pp.1389–1409.

Lakhssassi, N., Piya, S., Bekal, S., Liu, S., Zhou, Z., Bergounioux, C., Miao, L., Meksem, J., Lakhssassi, A., Jones, K., Kassem, M.A., Benhamed, M., Bendahmane, A., Lambert, K., Boualem, A., Hewezi, T. and Meksem, K. (2020). A pathogenesis-related protein GmPR08-Bet VI promotes a molecular interaction between the GmSHMT08 and GmSNAP18 in resistance to Heterodera glycines. *Plant Biotechnology Journal*, 18(8), pp.1810–1829.

Mair, A., Xu, S.-L., Branon, T.C., Ting, A.Y. and Bergmann, D.C. (2019). Proximity labeling of protein complexes and cell-type-specific organellar proteomes in Arabidopsis enabled by TurboID. *eLife*, 8.

Nonaka, H., Nakanishi, Y., Kuno, S., Ota, T., Mochidome, K., Saito, Y., Sugihara, F., Takakusagi, Y., Aoki, I., Nagatoishi, S., Tsumoto, K. and Sando, S. (2019). Design strategy for serine hydroxymethyltransferase probes based on retro-aldol-type reaction. *Nature Communications*, 10(1).

Otten, A.D., Sanders, M.M. and McKnight, G.S. (1988). The MMTV LTR Promoter is Induced by Progesterone and Dihydrotestosterone but not by Estrogen. *Molecular Endocrinology*, 2(2), pp.143–147.

Peremyslov, V.V., Morgun, E.A., Kurth, E.G., Makarova, K.S., Koonin, E.V. and Dolja, V.V. (2013). Identification of Myosin XI Receptors in Arabidopsis Defines a Distinct Class of Transport Vesicles. *The Plant Cell*, 25(8), pp.3022–3038.

Rancour, D.M., Park, S., Knight, S.D. and Bednarek, S.Y. (2004). Plant UBX Domain-containing Protein 1, PUX1, Regulates the Oligomeric Structure and Activity of Arabidopsis CDC48. *Journal of Biological Chemistry*, 279(52), pp.54264–54274.

Riggs, R., Wang, S., Singh, R. and Hymowitz, T. (1998). Possible Transfer of Resistance to Heterodera glycines from Glycine Tomentella to Glycine Max. *The Journal of Nematology*, 30(4S), pp.547–552.

Schroeder, M., Tsuchiya, T., He, S. and Eulgem, T. (2015). Use of enhancer trapping to identify pathogen-induced regulatory events spatially restricted to plant-microbe interaction sites. *Molecular Plant Pathology*, 17(3), pp.388–397.

Tisdale, E.J., Kelly, C. and Artalejo, C.R. (2004). Glyceraldehyde-3-phosphate Dehydrogenase Interacts with Rab2 and Plays an Essential Role in Endoplasmic Reticulum to Golgi Transport Exclusive of Its Glycolytic Activity. *Journal of Biological Chemistry*, 279(52), pp.54046–54052.

Tsuchiya, T. and Eulgem, T. (2013). An alternative polyadenylation mechanism coopted to the Arabidopsis RPP7 gene through intronic retrotransposon domestication. *Proceedings of the National Academy of Sciences*, 110(37), pp.E3535–E3543.

Vercollone, J.R., Balzar, M., Litvinov, S.V., Yang, W. and Cirulli, V. (2015). MMTV/LTR Promoter-Driven Transgenic Expression of EpCAM Leads to the Development of Large Pancreatic Islets. *Journal of Histochemistry & Cytochemistry*, 63(8), pp.613–625.

Wen, L., Chang, H.-X., Brown, P.J., Domier, L.L. and Hartman, G.L. (2019). Genome-wide association and genomic prediction identifies soybean cyst nematode resistance in common bean including a syntenic region to soybean Rhg1 locus. *Horticulture Research*, 6(1).

Wen, L., Yuan, C., Herman, T.K. and Hartman, G.L. (2017). Accessions of Perennial Glycine Species With Resistance to Multiple Types of Soybean Cyst Nematode (Heterodera glycines). *Plant Disease*, 101(7), pp.1201–1206.

Yang, L., Qin, L., Liu, G., Peremyslov, V.V., Dolja, V.V. and Wei, Y. (2014). Myosins XI modulate host cellular responses and penetration resistance to fungal pathogens. *Proceedings of the National Academy of Sciences*, 111(38), pp.13996–14001.

Zhang, H., Kjemtrup-Lovelace, S., Li, C., Luo, Y., Chen, L.P. and Song, B.-H. (2017). Comparative RNA-Seq Analysis Uncovers a Complex Regulatory Network for Soybean Cyst Nematode Resistance in Wild Soybean (Glycine soja). *Scientific Reports*, 7(1).

Zhang, Y., Song, G., Lal, N.K., Nagalakshmi, U., Li, Y., Zheng, W., Huang, P., Branon, T.C., Ting, A.Y., Walley, J.W. and Dinesh-Kumar, S.P. (2019). TurboID-based proximity labeling reveals that UBR7 is a regulator of N NLR immune receptor-mediated immunity. *Nature Communications*, 10(1).

Appendix 1: Cloning of HG type 0 nematode effectors and leveraging *in planta* assays for soybean cyst nematode (SCN) effector characterization

Derrick Grunwald, Andrew Bent

A1.1 Introduction

Soybean cyst nematode (SCN) pathology is largely dependent on the effector suite a nematode population uses, as well as whether the nematode encounters a resistant or susceptible cultivar (Mitchum et al., 2013). In the case of some resistant cultivars, a plant hypersensitive, programmed cell death (PCD)-like response occurs at the nematode feeding site resulting in nutrient deprivation and reduced nematode development (Endo, 1965). Furthermore, there have been multiple transcriptional studies that implicate differential transcriptional responses between resistant and susceptible cultivars underlying resistance (Khan et al., 2004; Klink et al., 2005; Klink et al., 2007a; Klink et al., 2007b; Klink et al., 2009; Klink et al., 2010; Kandoth et al., 2011). To date, there have been relatively few SCN effectors characterized and associated with the suppression of stress and defense responses during

feeding site development (Mitchum et al., 2013; Gardner et al., 2015; Noon et al., 2016; Yang et al., 2019; Wang et al., 2020). Moreover, even less is known about the interactions between SCN effectors and proteins encoded at the major resistance locus, *Rhg1* (Bekal et al., 2015).

SCN-plant studies have been greatly improved in recent years with the application of next-generation sequencing technologies. Low-abundance and non-stylet secreted putative effectors have been described using transcriptomic data, and such studies have greatly expanded the number of candidate effectors for molecular characterization (Gardner et al., 2018). Genomic studies have additionally expanded the number of candidate effectors through the discovery of promoter DOG-box elements (Eves van den Akker and Birch 2016; Eves-van den Akker et al., 2016; Masonbrink et al., 2019a). Further studies have additionally implicated roles of both transposons and alternative splicing in facilitating effector suite expansion (Masonbrink et al., 2019a; Masonbrink et al., 2019b). Altogether, such studies have provided potential targets for molecular characterization, and furnished a framework for cloning novel effectors from different nematode populations.

Below we describe cloning and initial characterization of a number of effectors derived from an HG type 0 nematode population for *in planta* expression in *Nicotiana benthamiana*. In particular, we discuss the use of effector overexpression constructs to test for suppression of cell death induced by α-SNAP$_{Rhg1}$, under the hypothesis that suppression is the result of either interaction with the protein itself or acting epistatically to PCD-like pathways. We further discuss the construction of plasmids which can be used to query binary interactions between *Rhg1*-encoded proteins and effectors (split-luciferase), as well as more global interaction

differences between *Rhg1*-encoded proteins during infection [proximity labeling (Mair et al., 2019; Zhang et al., 2019)].

A1.2 Results and discussion

We investigated the interaction between a number of SCN HG type 0 effectors and the α-SNAP$_{Rhg1}$ from *rhg1-a G. max*. Because the resistance responses to SCN mediated by the combination of *rhg1-a* and *Rhg4* are associated with an apparent PCD-like response and because there have been a number of SCN HG type 0 effectors found which can suppress PCD in yeast, we hypothesized that some of these effectors would be able to suppress the cell death induced by α-SNAP$_{Rhg1}$LC expression *in planta* (Bayless et al., 2016; Bayless et al., 2018; Wang et al., 2020). It is important to note, however, that the cell death resulting from α-SNAP$_{Rhg1}$ expression in *N. benthamiana* appears to be a delayed result of defective vesicle trafficking rather than a classic HR-type programmed cell death response. A second and simultaneous working hypothesis was that some SCN effectors, more likely those from HG type 2.5.7 or HG type 1.3.6 populations rather than HG type 0 effectors, act via more direct interaction with α-SNAPs or other vesicle trafficking components to suppress the *N. benthamiana* leaf cell death induced by α-SNAP$_{Rhg1}$ expression. To test these hypotheses, we co-expressed α-SNAP$_{Rhg1}$LC and the denoted effectors in *N. benthamiana* and assayed cell death visually as well as quantitively through electrolyte leakage.

Visually, expression of the α-SNAP$_{Rhg1}$LC was able to cause modest cell death when co-expressed with empty vector, and this cell death was generally less prevalent when co-expressed with α-SNAP$_{WT}$ (Figure 1A). The prevalence of cell death was variable when α-

$SNAP_{Rhg1}$LC was co-expressed with the SCN effectors. For example, while 33E05 and 34B08 did not alter α-$SNAP_{Rhg1}$LC-induced cell death, 2A05 and 2B10 were seemingly able to reduce the prevalence of induced cell death (Figure 1A). We further characterized cell death using electrolyte leakage experiments. Similar to visual inspection, there was less electrolyte leakage when α-$SNAP_{Rhg1}$LC was co-expressed with α-$SNAP_{WT}$ than with empty vector (Figure 1B). However, there was also some variability between assays. 33A09 and 32E03 eventually, by Day 7, produced greater electrolyte leakage than expression of α-$SNAP_{Rhg1}$LC alone (Figure 1B), possibly because they cause death of plant cells independent of α-$SNAP_{Rhg1}$LC. Follow-up studies would be required to test that hypothesis, but the prevalence of visually apparent cell death caused by 33A09 and 32E03 was only moderate (Figure 1A). Effectors 2A05 and 2B10 gave low cell death prevalence and low electrolyte leakage between the two assays.

Interestingly, 2A05 was one of the effectors described to suppress cell death induced by BAG6 in yeast (Wang et al., 2020). 2A05 is a protein of the venom allergen protein (VAP) class. VAPs from the beet cyst nematode have previously been shown to suppress basal immunity in *A. thaliana* mediated by damage-associated molecular pattern recognition (Lozano-Torres et al., 2014). Interestingly, some of the effectors described as suppressing PCD in yeast were not able to suppress cell death in *N. benthamiana* in our screen: 3H07, 34B08, 32E03 and 33A09 were not able to suppress cell death induced by α-$SNAP_{Rhg1}$LC while being able to suppress cell death from BAG6 in yeast (Wang et al., 2020). Future work should be directed to replicating the results of 2A05 and 2B10, as well as potentially screening additional HG type 0 effectors, with an emphasis on those that have previously been shown to suppress cell death in yeast,

e.g., VAPs. For the present Ph.D. thesis research, limited additional testing of 2A05 and 2B10 was carried out using a yeast assay system (Appendix 4).

The effectors used here are derived from an HG type 0 nematode population, which are avirulent on many commercially grown soybeans. We chose to work with this effector population due to both convenience and the potential of elucidating mechanisms that Hg type 0 nematode population induce disease on susceptible varieties. Future work should additionally be directed towards elucidating population-specific effectors that could be explanatory of virulence, with particular emphasis on HG type 2.5.7 nematode populations that are able to propagate on commercially relevant soybeans. These effectors may then be processed in the cell-death suppression pipeline as above or tested for direct physical interaction with *Rhg1*-encoded proteins using such methodologies as co-immunoprecipitation or proximity labeling. Effector candidates may be selected based on homology to known effectors and/or proximity to genomic features associated with nematode effectors (Eves van den Akker and Birch., 2016; Eves-van den Akker et al., 2016; Gardner et al., 2018; Masonbrink et al., 2019a; Masonbrink et al., 2019b).

The work we describe here will only show whether an effector is able to suppress cell death through acting in the same pathway of the α-SNAP$_{Rhg1}$-induced cell death. However, it remains an open question in the field whether there are effectors which directly interact with any of the *Rhg1*-encoded proteins, with particular interest in the α-SNAP, which is polymorphic between resistant and susceptible varieties. To begin to answer these questions, we have designed both split-luciferase and proximity labeling constructs. Split luciferase (Figure 4) will allow the interaction between a *Rhg1*-encoded protein and a putative effector to be

quantitatively measured *in planta*. Through this work, the HG type 0 effectors and *Rhg1*-encoded genes have been cloned into appropriate split-luciferase vectors. Complementary to this approach, proximity labeling (Figure 3) will allow characterization of protein complexes that form through the course of infection between *Rhg1*-encoded proteins, nematode effectors and additional *G. max* proteins. Here, the *Rhg1*-encoded genes and both NSF_{Ch07} and NSF_{RAN07} have been cloned into appropriate proximity labeling constructs. For both split-luciferase and proximity labeling experiments, an attractive starting point would be characterizing the interaction between putative SNARE-like protein effectors and the soybean vesicle trafficking components, with particular emphasis on α-SNAP and NSF (Bekal et al., 2015; Gardner et al., 2018).

A1.3 Methods

Source of Cloned Effectors The forty-seven effectors from HG type 0 nematode populations in pDonor/Zeo were gifted by the labs of Thomas Baum and Melissa Mitchum (Gao et al., 2001; Gao et al., 2003). These effectors are derived from cDNA libraries of gland-specific genes expressed during parasitism.

Vector Construction *in planta* effector constructs were made by first amplifying the open reading frame of Forrest or Williams 82 α-SNAPs derived from cDNA and cloning them into pDonor/Zeo by BP cloning per manufacturer's instruction (ThermoFisher). For transformation into the binary vector pSH174 (Figure 2), the open reading frame or the effector coding sequence was amplified using primers against the attL1 and attL2 sites in pDonor/Zeo by PCR

using the KapaHiFi polymerase (Kapa Biosystems). The amplicons were then purified using the Zymo Large Fragment DNA Recovery Kit (Zymo, Hercules CA), and assembled into the CaMV 35S promoter, nopaline synthase terminator containing binary vector pSH174 by ligase independent cloning (Huan et al., 1992). For construction of proximity labeling plasmids, *GmSNAP18*, *2580*, $NSF_{Ch07,}$ and NSF_{RAN07} were cloned into pMDC7-11, pMDC7-11 (Figure 3A), and pMDC7-9 (Figure 3B), respectively by LR cloning using manufacturer's instructions (ThermoFischer). Split-luciferase fusions were constructed as follows: amplicons from above were cloned into pJG014 or pLYY01 for N- and C-terminal fusions, respectively, by ligase independent cloning (Figure 4A and B).

Transient Expression in N. benthamiana *Nicotiana benthamiana* plants were used for agro-infiltration assays. Constructs derived from pSH174 were used to transform *Agrobacterium tumefaciens* strain GV3101 (pMP90) using the freeze-thaw method. Transformed cells were cultured in LB with both kanamycin and rifampicin at concentrations of 25 μg/mL and 100 μg/mL, respectively. Cultured cells were harvested by centrifugation and resuspended in MMA induction buffer [1 L MMA: 5 g MS salts, 1.95 g MES, 20 g sucrose, 200 mM acetosyringone, pH = 5.6]. Suspensions were infiltrated into *N. benthamian*a using a tuberculin syringe with no needle. For cell death suppression assays, a total of OD_{600} = 0.6 was co-infiltrated into leaves at a 1:1 ratio of effector to α-$SNAP_{Rhg1}$LC and tissue collected two days after infiltration for electrolyte leakage assays.

Electrolyte leakage assay Assays were performed essentially as in Kessens et al., 2018. In brief, approximately 24 hours after infiltration, six to eight leaf punches were taken from two different leaves on the same plant and placed in a single well of a 12 well plate, representing a single biological replicate. Leaf discs were washed for 30 minutes with sterile deionized water while rotating at 50 rpm. After 30 minutes, water was removed and replaced with 4 mL of fresh water and conductivity measured on an ECTestr 11+ MultiRange conductivity meter (Oakton), representing the 24 hr timepoint. Subsequent timepoints were performed as above.

Primers Used in Appendix

LIC_FW_KOZAK: CGACGACAAGACCGTGACCATGTACAAAAAAGCAGGCTAT

LIC_REV_STOP: GAGGAGAAGAGCCGTCAAACTTTGTACAAGAAAGC

2580_AttB_FW: GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGTCTCCGGCCGCCGGA

2580_AttB_REV:

GGGGACCACTTTGTACAAGAAAGCTGGGTTTTATGACTTGCTACTAAAAGCATTATATATGTTGGTGG

WM_AttB_FW : GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGCCGATCAGTTATCGAAGGG

WM_AttB_REV:

GGGGACCACTTTGTACAAGAAAGCTGGGTTTTAAGTAAGATCATCCTCCTCAAGTTCTTTGG

NSF07_AttB_Fw:GGGGACAAGTTTGTACAAAAAAGCAGGCTTCACCATGGCGAGTCGGTTCGGGTTAT

CG

NSF07_AttB_REV: GGGGACCACTTTGTACAAGAAAGCTGGGTGTAACCTAACAACATCCTGGAGGCA

RAN_AttB_FW: GGGGACAAGTTTGTACAAAAAAGCAGGCTCGGCGAGTCAGTTCGGGTTATCGTCT

RAN_AttB_REV:

GGGGACCACTTTGTACAAGAAAGCTGGGTCTCATAACCTAACAACATCCTGGAGGCAATCA

LC_AttB_REV: GGGGACCACTTTGTACAAGAAAGCTGGGTTCAAGTAATAACCTCATACTCCTCAAG

A1.4 References

Bayless, A.M., Smith, J.M., Song, J., McMinn, P.H., Teillet, A., August, B.K. and Bent, A.F. (2016). Disease resistance through impairment of α-SNAP–NSF interaction and vesicular trafficking by soybean Rhg1. *Proceedings of the National Academy of Sciences*, 113(47), pp.E7375–E7382.

Bayless, A.M., Zapotocny, R.W., Grunwald, D.J., Amundson, K.K., Diers, B.W. and Bent, A.F. (2018). An atypical N-ethylmaleimide sensitive factor enables the viability of nematode-resistant Rhg1 soybeans. *Proceedings of the National Academy of Sciences*, 115(19), pp.E4512–E4521.

Bekal, S., Domier, L.L., Gonfa, B., Lakhssassi, N., Meksem, K. and Lambert, K.N. (2015). A SNARE-Like Protein and Biotin Are Implicated in Soybean Cyst Nematode Virulence. *PLOS ONE*, 10(12), p.e0145601.

Endo, B. (1965). Histological responses of resistant and susceptible soybean varieties, and backcross progeny to entry and development of Heterodera glycines. *Phytopathology*, 55, pp.375–381.

Eves-van den Akker, S., Laetsch, D.R., Thorpe, P., Lilley, C.J., Danchin, E.G.J., Da Rocha, M., Rancurel, C., Holroyd, N.E., Cotton, J.A., Szitenberg, A., Grenier, E., Montarry, J., Mimee, B., Duceppe, M.-O., Boyes, I., Marvin, J.M.C., Jones, L.M., Yusup, H.B., Lafond-Lapalme, J., Esquibet, M., Sabeh, M., Rott, M., Overmars, H., Finkers-Tomczak, A., Smant, G., Koutsovoulos, G., Blok, V., Mantelin, S., Cock, P.J.A., Phillips, W., Henrissat, B., Urwin, P.E., Blaxter, M. and Jones, J.T. (2016). The genome of the yellow potato cyst nematode, Globodera rostochiensis, reveals insights into the basis of parasitism and virulence. *Genome Biology*, 17(1).

Eves-van den Akker, S. and Birch, P.R.J. (2016). Opening the Effector Protein Toolbox for Plant–Parasitic Cyst Nematode Interactions. *Molecular Plant*, 9(11), pp.1451–1453.

Gardner, M., Dhroso, A., Johnson, N., Davis, E.L., Baum, T.J., Korkin, D. and Mitchum, M.G. (2018). Novel global effector mining from the transcriptome of early life stages of the soybean cyst nematode Heterodera glycines. *Scientific Reports*, 8(1).

Huan, R., Serventi, I. and Moss, J. (1992). Rapid, reliable ligation-independent cloning of PCR products using modified plasmid vectors. *Biotechniques*, 13(4), pp.515–518.

Kandoth, P.K., Ithal, N., Recknor, J., Maier, T., Nettleton, D., Baum, T.J. and Mitchum, M.G. (2011). The Soybean Rhg1 Locus for Resistance to the Soybean Cyst Nematode Heterodera glycines Regulates the Expression of a Large Number of Stress- and Defense-Related Genes in Degenerating Feeding Cells. *Plant Physiology*, 155(4), pp.1960–1975.

Kessens, R., Sorensen, N. and Kabbage, M. (2018). An inhibitor of apoptosis (SfIAP) interacts with SQUAMOSA promoter-binding protein (SBP) transcription factors that exhibit pro-cell death characteristics. *Plant Direct*, 2(8), p.e00081.

Khan, R., Alkharouf, N., MacDonald, M., Chouikha, I., Meyer, S., Grefenstette, J., Knap, H. and Matthews, B. (2004). Microarray Analysis of Gene Expression in Soybean Roots Susceptible to the Soybean Cyst Nematode Two Days Post Invasion. *The Journal of Nematology*, 36(3), pp.214–248.

Klink, V.P., Alkharouf, N., MacDonald, M. and Matthews, B. (2005). Laser Capture Microdissection (LCM) and Expression Analyses of Glycine max (Soybean) Syncytium Containing Root Regions Formed by the Plant Pathogen Heterodera glycines (Soybean Cyst Nematode). *Plant Molecular Biology*, 59(6), pp.965–979.

Klink, V.P., Hosseini, P., Matsye, P., Alkharouf, N.W. and Matthews, B.F. (2009). A gene expression analysis of syncytia laser microdissected from the roots of the Glycine max (soybean) genotype PI 548402 (Peking) undergoing a resistant reaction after infection by Heterodera glycines (soybean cyst nematode). *Plant Molecular Biology*, 71(6), pp.525–567.

Klink, V.P., Hosseini, P., Matsye, P.D., Alkharouf, N.W. and Matthews, B.F. (2010). Syncytium gene expression in Glycine max[PI 88788] roots undergoing a resistant reaction to the parasitic nematode Heterodera glycines. *Plant Physiology and Biochemistry*, 48(2–3), pp.176–193.

Klink, V.P., Overall, C.C., Alkharouf, N.W., MacDonald, M.H. and Matthews, B.F. (2007a). A time-course comparative microarray analysis of an incompatible and compatible response by Glycine max (soybean) to Heterodera glycines (soybean cyst nematode) infection. *Planta*, 226(6), pp.1423–1447.

Klink, V.P., Overall, C.C., Alkharouf, N.W., MacDonald, M.H. and Matthews, B.F. (2007b). 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(6), pp.1389–1409.

Lozano-Torres, J.L., Wilbers, R.H.P., Warmerdam, S., Finkers-Tomczak, A., Diaz-Granados, A., van Schaik, C.C., Helder, J., Bakker, J., Goverse, A., Schots, A. and Smant, G. (2014). Apoplastic Venom Allergen-like Proteins of Cyst Nematodes Modulate the Activation of Basal Plant Innate Immunity by Cell Surface Receptors. *PLoS Pathogens*, 10(12), p.e1004569.

Mair, A., Xu, S.-L., Branon, T.C., Ting, A.Y. and Bergmann, D.C. (2019). Proximity labeling of protein complexes and cell-type-specific organellar proteomes in Arabidopsis enabled by TurboID. *eLife*, 8.

Masonbrink, R., Maier, T.R., Muppirala, U., Seetharam, A.S., Lord, E., Juvale, P.S., Schmutz, J., Johnson, N.T., Korkin, D., Mitchum, M.G., Mimee, B., den Akker, S.E., Hudson, M., Severin, A.J. and Baum, T.J. (2019a). The genome of the soybean cyst nematode (Heterodera glycines) reveals complex patterns of duplications involved in the evolution of parasitism genes. *BMC Genomics*, 20(1).

Masonbrink, R., Maier, T.R., Seetharam, A.S., Juvale, P.S., Baber, L., Baum, T.J. and Severin, A.J. (2019b). SCNBase: a genomics portal for the soybean cyst nematode (Heterodera glycines). *Database*, 2019.

Mitchum, M.G., Hussey, R.S., Baum, T.J., Wang, X., Elling, A.A., Wubben, M. and Davis, E.L. (2013). Nematode effector proteins: an emerging paradigm of parasitism. *New Phytologist*, 199(4), pp.879–894.

Noon, J.B., Qi, M., Sill, D.N., Muppirala, U., Eves-van den Akker, S., Maier, T.R., Dobbs, D., Mitchum, M.G., Hewezi, T. and Baum, T.J. (2016). APlasmodium-like virulence effector of the soybean cyst nematode suppresses plant innate immunity. *New Phytologist*, 212(2), pp.444–460.

Wang, J., Yeckel, G., Kandoth, P.K., Wasala, L., Hussey, R.S., Davis, E.L., Baum, T.J. and Mitchum, M.G. (2020). Targeted suppression of soybean BAG6-induced cell death in yeast by soybean cyst nematode effectors. *Molecular Plant Pathology*, 21(9), pp.1227–1239.

Yang, S., Dai, Y., Chen, Y., Yang, J., Yang, D., Liu, Q. and Jian, H. (2019). A Novel G16B09-Like Effector From Heterodera avenae Suppresses Plant Defenses and Promotes Parasitism. *Frontiers in Plant Science*, 10.

Zhang, Y., Song, G., Lal, N.K., Nagalakshmi, U., Li, Y., Zheng, W., Huang, P., Branon, T.C., Ting, A.Y., Walley, J.W. and Dinesh-Kumar, S.P. (2019). TurboID-based proximity labeling reveals that UBR7 is a regulator of N NLR immune receptor-mediated immunity. *Nature Communications*, 10(1).

A1.5 Figures

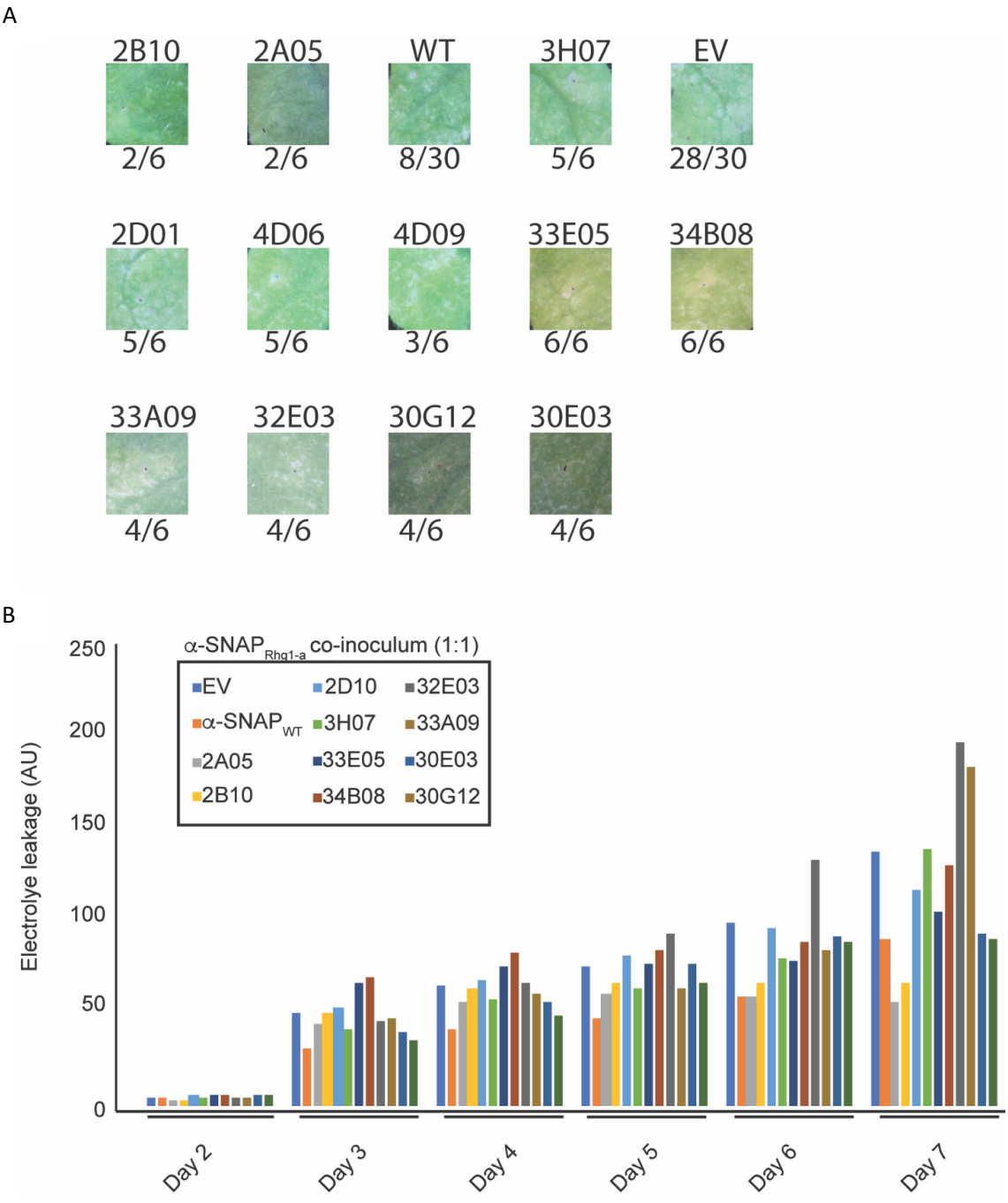


Figure 1: (A) Representative leaf images of cell death resultant from co-expression of α-

$SNAP_{Rhg1}$LC and denoted effector, empty vector or α-$SNAP_{WT}$. All co-inoculations are a 1:1 ratio with a total OD600 of 0.6. The ratio of leaves with the cell death phenotype to the total number inoculated is represented under the image. (B) Cell death in A quantified using an electrolyte leakage assay. Each bar represents a total of four leaves from two biological replicates.

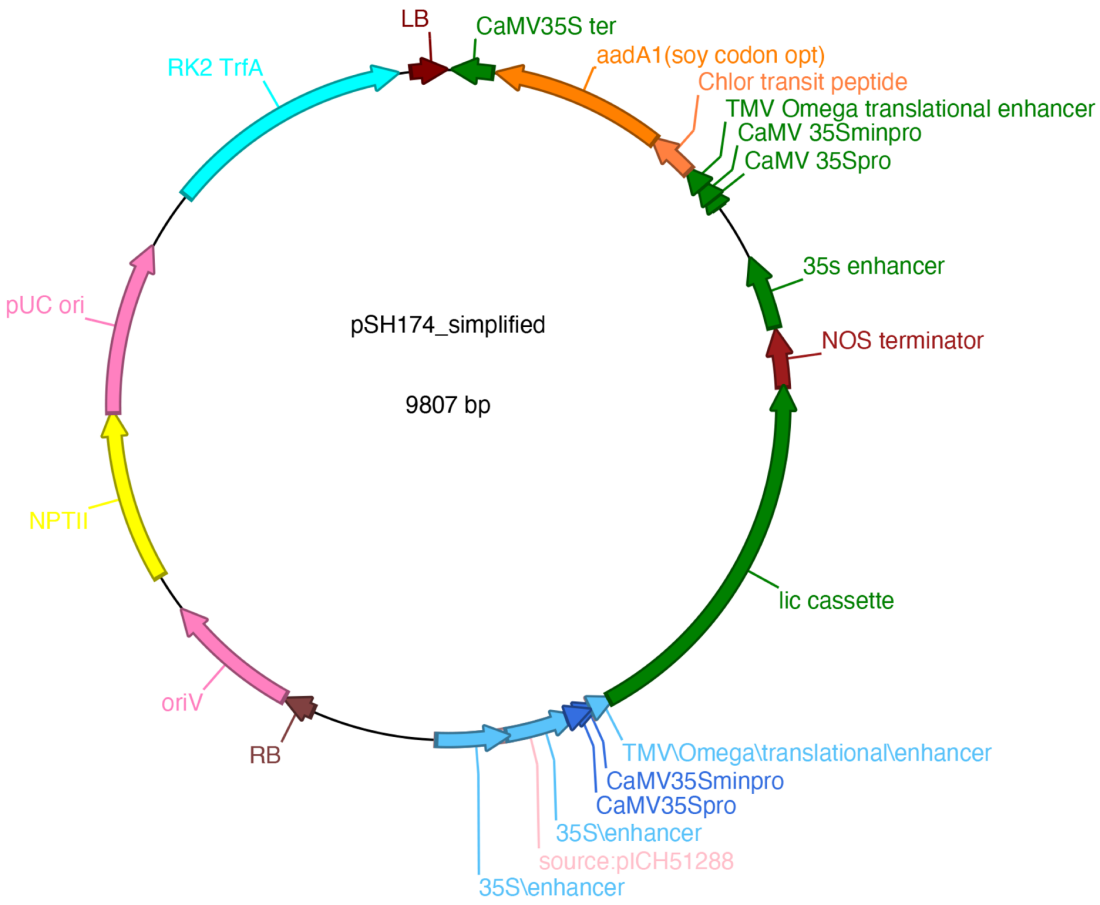
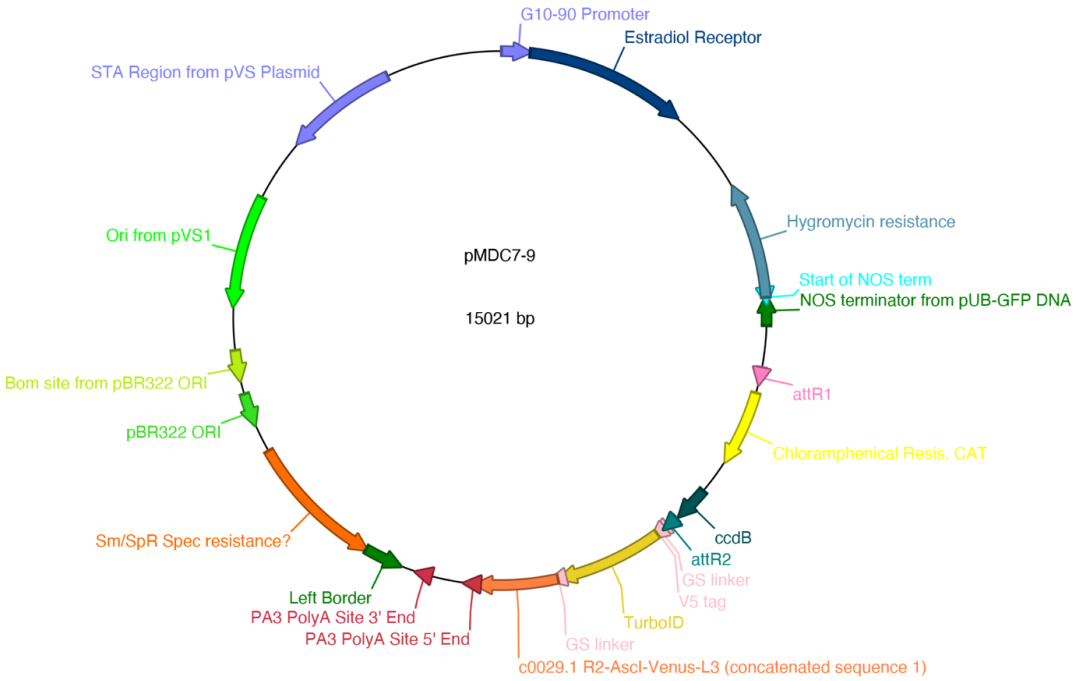


Figure 2: The pSH174 binary plasmid map. This vector was used for cloning effectors or α-SNAP by ligase independent cloning for expression in *N. benthamiana*. Bacterial selection is kanamycin, and plant selection is spectinomycin.

A


G10-90 Promoter
Estradiol Receptor
STA Region from pVS Plasmid
Hygromycin resistance
Ori from pVS1
pMDC7-9
15021 bp
Start of NOS term
NOS terminator from pUB-GFP DNA
Bom site from pBR322 ORI
attR1
pBR322 ORI
Chloramphenical Resis, CAT
ccdB
attR2
Sm/SpR Spec resistance?
GS linker
V5 tag
Left Border
TurboID
PA3 PolyA Site 3' End
GS linker
PA3 PolyA Site 5' End
c0029.1 R2-AscI-Venus-L3 (concatenated sequence 1)

B

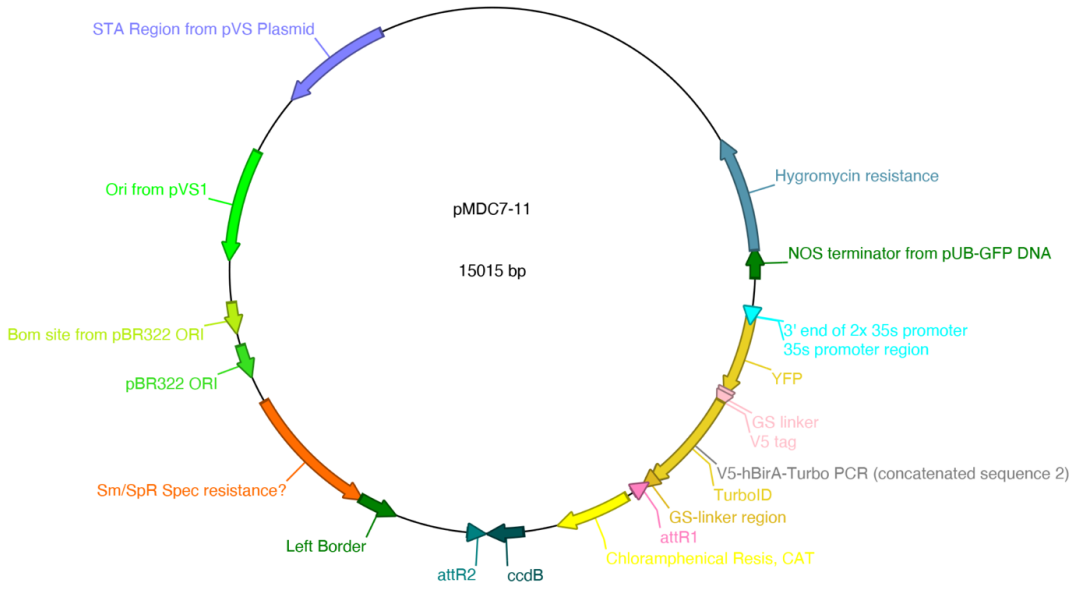
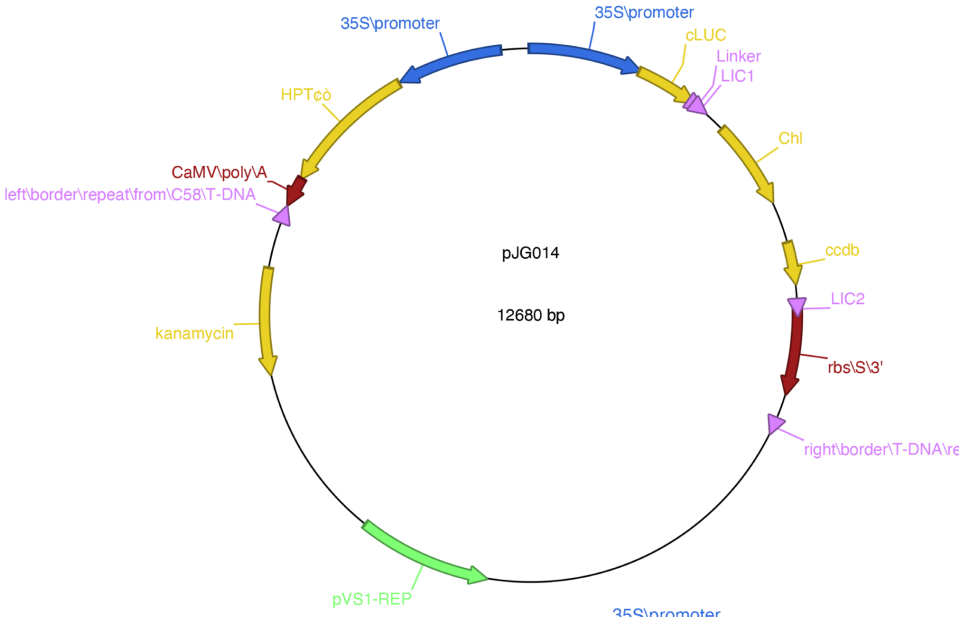

STA Region from pVS Plasmid
Hygromycin resistance
Ori from pVS1
pMDC7-11
15015 bp
NOS terminator from pUB-GFP DNA
3' end of 2x 35s promoter
35s promoter region
Bom site from pBR322 ORI
YFP
pBR322 ORI
GS linker
V5 tag
V5-hBirA-Turbo PCR (concatenated sequence 2)
Sm/SpR Spec resistance?
TurboID
GS-linker region
Left Border
attR1
Chloramphenical Resis, CAT
attR2
ccdB

Figure 3: The pMDC7-9 (A) and pMDC7-11 (B) binary vectors for cloning effectors or *Rhg1*-encoded proteins with a C- or N-terminally fused biotin ligase for proximity labeling experiments. Cloning is done by Gateway. Bacterial selection is spectinomycin, and plant selection is hygromycin.

A


35S\promoter
35S\promoter
cLUC
Linker
LIC1
HPTçö
Chl
CaMV\poly\A
left\border\repeat\from\C58\T-DNA
pJG014
12680 bp
ccdb
LIC2
rbs\S\3'
kanamycin
right\border\T-DNA\re
pVS1-REP

B

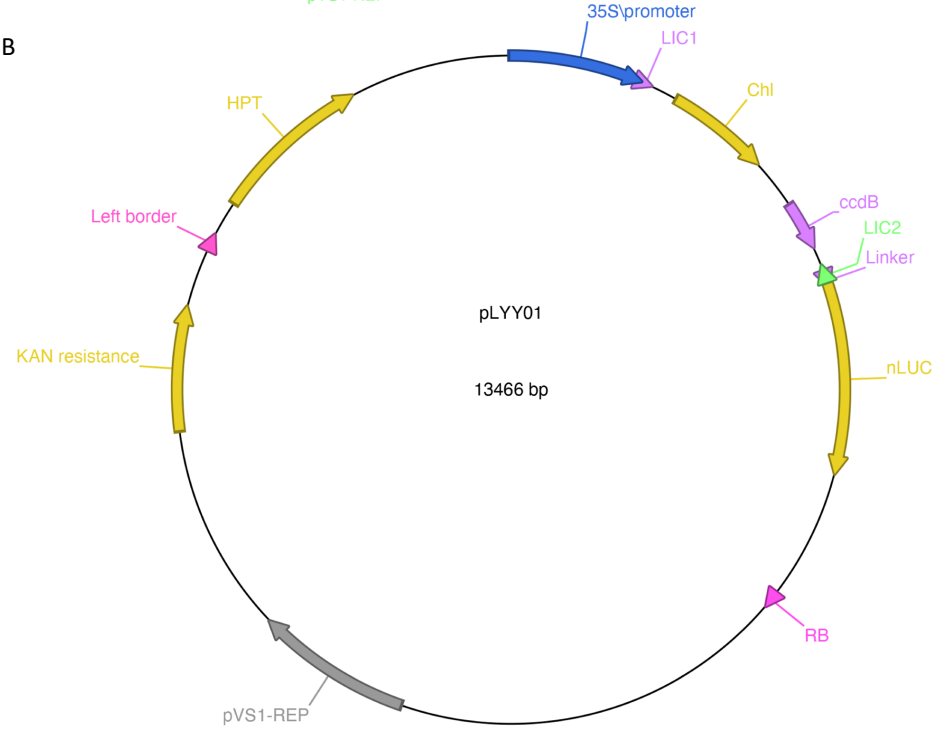

35S\promoter
LIC1
HPT
Chl
Left border
ccdB
LIC2
Linker
pLYY01
KAN resistance
13466 bp
nLUC
RB
pVS1-REP

Figure 4: The pJG014 (A) and pLYY01 (B) binary vectors for cloning effectors or *Rhg1*-encoded proteins with a N-terminally fused C-luciferase or C-terminally fused N-luciferase for luciferase complementation experiments ("split luciferin" assays). Cloning is done by ligase independent cloning. Bacterial selection is kanamycin, and plant selection is hygromycin.

Appendix 2: Design of α-SNAP constructs responsive to other biotrophic pathogens, most particularly *Hyaloperonospora arabidopsidis*, HPA.

Derrick Grunwald, Yulin Du, Andrew Bent

I performed all cloning of the pro*MyoB1*::*GmSNAP18* constructs. *Arabidopsis* was transformed and T1 seeds were screened by Yulin Du.

A2.1 Introduction

Both biotrophic and hemibiotrophic pathogens constitute a wide swath of plant pathogens and includes such widely known pathogens as *Hyaloperonospora arabidopsidis* (the causal agent of downy mildew, HPA), *Pseudomonas syringae* (causal agent of bacterial speck), and soybean cyst nematode (SCN). In general, within the spectrum of biotrophic interactions, plants and pathogens exchange nutrients and other small molecules to ultimately benefit the pathogen's growth and reproduction. In order to survive in a plant, many biotrophic pathogens have evolved effector suites that are biologically active against plant hosts (reviewed in Fabro

et al., 2011; Mitchum et al., 2013; Spanu and Panstruga, 2017; Dillon et al., 2019). Recognition of either effectors themselves or the products of the biological activity of the effectors forms the basis of recognition and resistance in incompatible plant-microbe interactions (Jones and Dangl, 2006; van der Hoorn and Kamoun, 2008). However, qualitative resistance: resistance inherited through so called resistance (R) genes in a mendelian fashion, is not available for all plant pathogens and is often ecotype-dependent, e.g., in HPA (Steinbrenner et al., 2015; Goritschnig et al., 2016). Therefore, manipulating the outputs of an otherwise compatible interaction is an attractive method of introducing disease resistance in genetically tractable plant species.

Executor genes, as they occur naturally, are a family of genes that mediate programmed cell death in response to pathogen infection, and include such well known proteins as Xa10, Xa23, Xa27 (from rice) as well as BS3 and BS4C-R (from pepper) (Gu et al., 2005; Römer et al., 2007; Strauss et al., 2012; Tian et al., 2014; Wang et al., 2015). In their native systems, these executor genes are regulated by promoters that have TALE effector (produced by *Xanthomonas* species) recognition sequence (Strauß et al., 2012). Artificial integration of executor genes is becoming an increasingly feasible method to establish disease resistance in cisgenic systems, with notable success in the wheat/*X. oryzae* pv. oryzicola pathosystem (Hummel et al., 2012). For transgenic systems, the cloning of executor genes or genes that can mediate programmed cell death is theoretically possible for any plant-pathogen interaction with sufficient knowledge of genes that are specifically regulated by the pathogen.

Recent microarray and enhancer trapping experiments have helped to elucidate genes specifically regulated by virus, bacteria, oomycetes and nematode plant pathogens (Fridborg et

al., 2004; Kandoth et al., 2011; Schroeder et al., 2016; Zhang et al., 2017). One such gene regulated by oomycetes is the *Myosin Binding Protein 1* (*MyoB1*; *At1g08800*). MyoB1 is a protein that acts as a receptor for type XI myosin, themselves important in mediating a unique vesicle trafficking pathway, seemingly central to penetration resistance in *A. thaliana* (Peremyslov et al., 2013; Yang et al., 2014). In enhancer trapping experiments, *MyoB1* transcript was increased specifically in response to HPA infection at haustoria (Schroeder et al., 2016). Here, we propose utilizing the α-SNAP from *rhg1-a* soybeans as an executor gene to confer resistance to HPA in *A. thaliana.* Below, we describe cloning and generating T2 *A. thaliana* containing pro*MyoB1*::*GmSNAP18* constructs.

A2.2 Methods

Vector Construction *MyoB1* promoter fusion constructs were made by amplifying the 2.3kb region upstream of the first exon of *Arabidopsis thaliana* Col-0 *MyoB1*, the nopaline synthase terminator from pSM101, and the open reading frame of soybean *GmSNAP18* from cDNAs generated from Forrest (α-SNAP$_{Rhg1}$LC) or Williams 82 (α-SNAP$_{WT}$) mRNA through PCR using KapaHiFi polymerase (Kapa biosystems). PCR amplicons were purified using the Zymo Large Fragment DNA Recovery Kit (Zymo, Hercules CA) and assembled into pBlueScript through Gibson Assembly (Figure 1). The promoter-α-SNAP-nosT cassette was then amplified using primers with AscI/XbaI restriction sites and digested with these enzymes (New England Biosciences). The digested fragment was purified using the Zymo Large Fragment DNA Recovery Kit (Zymo, Hercules CA), and ligated into the binary vector pSM101 using T4 DNA ligase (New England Biosciences) to form plasmids shown in Figure 2.

Arabidopsis Transformation *A. thaliana* was transformed using floral dip technique (Clough and Bent, 1998). In brief *Agrobacterium tumefaciens* GV3101 (pMP90) harboring pSM101-EV, pSM101-pro*MyoB1*::α-SNAP$_{Rhg1}$LC and pSM101-pro*MyoB1*::α-SNAP$_{WT}$ was grown overnight in LB+10 mM morpholinoethanesulfonic acid (MES) + 20 μM acetosyringone + 25 μg/mL kanamycin + 100 μg/mL rifamycin. Cells were centrifuged at 4500 rpm and resuspended in 5% sucrose solution to a final OD_{600} of 0.8. Silwet L-77 was added to a final concentration of 0.05%. Flower buds of 5 to7-week-old *A. thaliana* was then dipped into bacterial solution for 10 seconds, and excess solution removed by drying with kimwipe paper. The process was repeated for about 12 plants per construct for each of the three constructs. The transformed *A. thaliana* was grown in the dark for 16 hours before being returned to normal growth conditions (22°C, 130-150 μE m^{-2} s^{-1}; 16-hour light).

Hygromycin selection ½ MS + sucrose media containing hygromycin was prepared (1L: 2.165 Murashige and Skoog [MS] salts, 10 g sucrose, 12 g agar, 25 μg/mL hygromycin). Approximately 500 seeds, resuspended in 0.1% agarose, were spread onto plates. Plates were then stored in the cold for 2 days, before being allowed to sit on bench for 8 hours. Plates were then transferred to the dark for 4 to 5 days. Transformants were selected based on hypocotyl elongation and grown in soil to produce T2 seeds.

Primers used in the appendix:

MyoB1-NosT-SNAP-FW: CGTCTGCAGTAAAATTACAGATGGCCGATCAGTTATCG

MyoB1-NosT-WT-RV: GCCAAATGTTTGAACGATCGTCAAGTAAGATCATCCTCC

MyoB1-NosT-LC-RV: GCCAAATGTTTGAACGATCGTCAAGTAATAACCTCATACTCC

MyoB1-Promoter-XbaI-FW:

NNNNNNNNNNTCTAGATAATTTTATATACACAAATAAAGAAACCAGAAAGAC

MyoB1-NosT-AscI-RV: NNNNNNNNNNGGCGCGCCGATCTAGTAACATAGATGACACCG

MyoB1-NosT-Gibson-RV: TGAACGATCGCTGTAATTTTACTGCAGACG

NosT-MyoB1-Gibson-FW: AAAATTACAGCGATCGTTCAAACATTTG

NosT-MyoB1-Gibson-RV: TGGCGGCCGCGATCTAGTAACATAGATGACAC

MyoB1-pBS-Gibson-FW: TTACTAGATCGCGGCCGCCACCGCGGTG

A2.3 References:

Clough, S.J. and Bent, A.F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *The Plant Journal*, 16(6), pp.735–743.

Dillon, M.M., Almeida, R.N.D., Laflamme, B., Martel, A., Weir, B.S., Desveaux, D. and Guttman, D.S. (2019). Molecular Evolution of Pseudomonas syringae Type III Secreted Effector Proteins. *Frontiers in Plant Science*, 10.

Fabro, G., Steinbrenner, J., Coates, M., Ishaque, N., Baxter, L., Studholme, D.J., Körner, E., Allen, R.L., Piquerez, S.J.M., Rougon-Cardoso, A., Greenshields, D., Lei, R., Badel, J.L., Caillaud, M.-C., Sohn, K.-H., Van den Ackerveken, G., Parker, J.E., Beynon, J. and Jones, J.D.G. (2011). Multiple Candidate Effectors from the Oomycete Pathogen Hyaloperonospora arabidopsidis Suppress Host Plant Immunity. *PLoS Pathogens*, 7(11), p.e1002348.

Goritschnig, S., Steinbrenner, A.D., Grunwald, D.J. and Staskawicz, B.J. (2016). Structurally distinct Arabidopsis thaliana NLR immune receptors recognize tandem WY domains of an oomycete effector. *New Phytologist*, 210(3), pp.984–996.

Gu, K., Yang, B., Tian, D., Wu, L., Wang, D., Sreekala, C., Yang, F., Chu, Z., Wang, G.-L., White, F.F. and Yin, Z. (2005). R gene expression induced by a type-III effector triggers disease resistance in rice. *Nature*, 435(7045), pp.1122–1125.

van der Hoorn, R.A.L. and Kamoun, S. (2008). From Guard to Decoy: A New Model for Perception of Plant Pathogen Effectors. *The Plant Cell*, 20(8), pp.2009–2017.

Hummel, A.W., Doyle, E.L. and Bogdanove, A.J. (2012). Addition of transcription activator-like effector binding sites to a pathogen strain-specific rice bacterial blight resistance gene makes it effective against additional strains and against bacterial leaf streak. *New Phytologist*, 195(4), pp.883–893.

Jones, J.D.G. and Dangl, J.L. (2006). The plant immune system. *Nature*, [online] 444(7117), pp.323–329. Available at: https://www.nature.com/articles/nature05286.

Mitchum, M.G., Hussey, R.S., Baum, T.J., Wang, X., Elling, A.A., Wubben, M. and Davis, E.L. (2013). Nematode effector proteins: an emerging paradigm of parasitism. *New Phytologist*, 199(4), pp.879–894.

Peremyslov, V.V., Morgun, E.A., Kurth, E.G., Makarova, K.S., Koonin, E.V. and Dolja, V.V. (2013). Identification of Myosin XI Receptors in Arabidopsis Defines a Distinct Class of Transport Vesicles. *The Plant Cell*, 25(8), pp.3022–3038.

Römer, P., Hahn, S., Jordan, T., Strauß, T., Bonas, U. and Lahaye, T. (2007). Plant Pathogen Recognition Mediated by Promoter Activation of the Pepper Bs3 Resistance Gene. *Science*, 318(5850), pp.645–648.

Spanu, P.D. and Panstruga, R. (2017). Editorial: Biotrophic Plant-Microbe Interactions. *Frontiers in Plant Science*, 8.

Steinbrenner, A.D., Goritschnig, S. and Staskawicz, B.J. (2015). Recognition and Activation Domains Contribute to Allele-Specific Responses of an Arabidopsis NLR Receptor to an Oomycete Effector Protein. *PLOS Pathogens*, 11(2), p.e1004665.

Strauss, T., van Poecke, R.M.P., Strauss, A., Romer, P., Minsavage, G.V., Singh, S., Wolf, C., Strauss, A., Kim, S., Lee, H.-A., Yeom, S.-I., Parniske, M., Stall, R.E., Jones, J.B., Choi, D., Prins, M. and Lahaye, T. (2012). RNA-seq pinpoints a Xanthomonas TAL-effector activated resistance gene in a large-crop genome. *Proceedings of the National Academy of Sciences*, 109(47), pp.19480–19485.

Tian, D., Wang, J., Zeng, X., Gu, K., Qiu, C., Yang, X., Zhou, Z., Goh, M., Luo, Y., Murata-Hori, M., White, F.F. and Yin, Z. (2014). The Rice TAL Effector–Dependent Resistance Protein XA10 Triggers Cell Death and Calcium Depletion in the Endoplasmic Reticulum. *The Plant Cell*, 26(1), pp.497–515.

Wang, C., Zhang, X., Fan, Y., Gao, Y., Zhu, Q., Zheng, C., Qin, T., Li, Y., Che, J., Zhang, M., Yang, B., Liu, Y. and Zhao, K. (2015). XA23 Is an Executor R Protein and Confers Broad-Spectrum Disease Resistance in Rice. *Molecular Plant*, 8(2), pp.290–302.

Yang, L., Qin, L., Liu, G., Peremyslov, V.V., Dolja, V.V. and Wei, Y. (2014). Myosins XI modulate host cellular responses and penetration resistance to fungal pathogens. *Proceedings of the National Academy of Sciences*, 111(38), pp.13996–14001.

A2.4 Figures

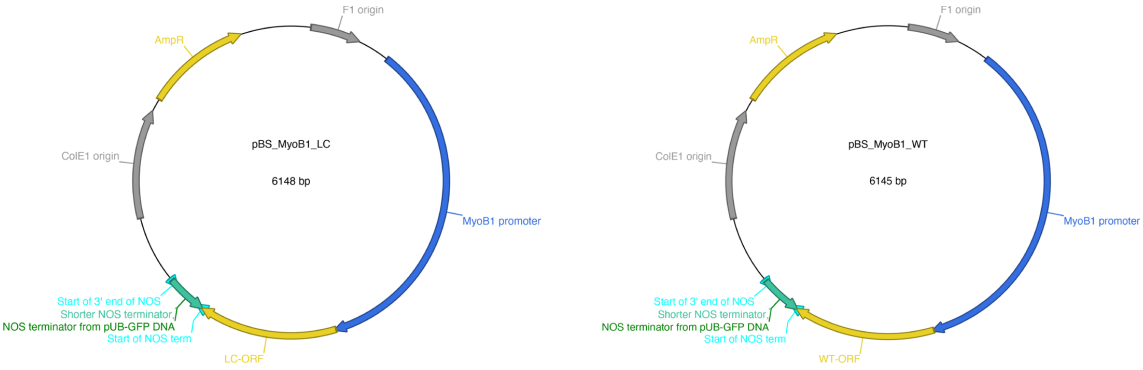


Figure 1: pBluescript plasmid maps generated for subcloning the α-SNAP driven by the *A. thaliana MyoB1* promoter into pSM101. Bacterial selection is ampicillin.

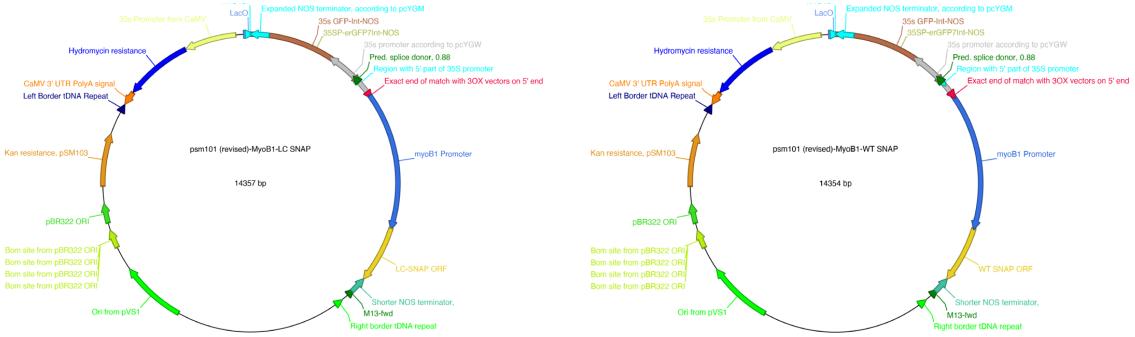


Figure 2: Plant binary vector pSM101 with *A. thaliana MyoB1* promoter driving expression of α-$SNAP_{Rhg1}$LC (LC) or α-$SNAP_{WT}$ (WT). Plants may be selected for hygromycin resistance or screened by GFP expression. Bacterial selection is kanamycin.

Appendix 3: Design and preliminary characterization of novel α-SNAPs with point mutations at the NSF interacting surface.

Derrick Grunwald, Huarui Zhou, Abbie Rogers, Andrew Bent

Andrew and I conceived of mutations. I performed cloning and preliminary characterization in *N. benthamiana* with Huraui Zhou. Preliminary nematode demographic assays were performed by Abbie Rogers.

A3.1 Introduction

Of the genes and gene products from the *Rhg1* locus, the best studied is the α-SNAP encoded by *Glyma.18g022500*. Work from this lab and others have demonstrated that the α-SNAP gene at the *Rhg1* locus encodes a protein that has amino acid polymorphisms distinct between *rhg1-a*, *rhg1-b*, and now *rhg1-cs* haplotypes, and these polymorphisms have functional consequences for the protein's ability to interact with the NSF ATPase (Cook et al., 2012; Cook et al., 2014; Lee et al., 2015; Bayless et al., 2016; Bayless et al., 2018). In addition to NSF, the α-SNAP at *Rhg1* has been shown to interact with the $SHMT_{Rhg4}$ in *rhg1-a* haplotypes, a pathogenesis related protein (Pr08-BetVI), t-SNAREs, and even an SCN-encoded SNARE-like

protein (Bekal et al., 2015; Lakhssassi et al., 2020; Dong et al., 2020). Broadly speaking, α-SNAP has the potential to interact with myriad components of both canonical and noncanonical pathways, vesicle trafficking or otherwise, and including the above but also such well characterized proteins as Munc18, Sec1, and various components of the HOPS complex (Lobingier et al., 2014; Morey et al., 2017; Song et al., 2017; Stepien et al., 2019). Each of these interactions represent a potential leverage point for functional modification through systematic mutagenesis.

Systematic and site-directed mutagenesis is a common method used to test the relative contributions of select amino acids on a protein's overall function. These methodologies have been extensively used and have even found application in probing the functions of α-SNAP and $SHMT_{Rhg4}$ (Marz et al., 2003; Kandoth et al., 2017). In the context of α-SNAP, systematic mutagenesis of mammalian α-SNAP uncovered distinct sets of residues that could either enhance SNARE binding, SNARE disassembly, both or neither (Marz et al., 2003). Additionally, systematic mutagenesis of the $SHMT_{Rhg4}$ elucidated residues critical for structural stability, ligand binding, enzyme kinetics and protein interaction sites, including putative sites of interaction with α-$SNAP_{Rhg1}$ (Kandoth et al., 2017). In addition to site-directed mutagenesis methods, random mutagenesis with chemical mutagens followed by TILLING, use of mutator strains, use of CRISPR-Cas9 methods, or leveraging of natural populations are all methods of discriminating protein domain function and/or creating novel alleles of a protein of interest (McCallum et al., 2000; Muteeb and Sen, 2010; Ford et al., 2018).

Here, we describe use of site directed mutagenesis to both probe the function of α-$SNAP_{Rhg1}$ and create novel alleles specifically at the C-terminal region. We focused on the C-

terminal region due to its existing variability in natural populations, and because it is known to be the site of interaction with NSF (Barnard et al., 1996; Cook et al., 2014; Lee et al., 2015; Zhao et al., 2015a; Bayless et al., 2016). We describe construction of 28 novel alleles and show that all tested alleles are at least partially functional in that they are able to cause cell death when transiently expressed in *N. benthamiana*. We further describe preliminary results of the resistance-conferring capacity of two of the alleles when expressed in a transgenic, detached root experiment. These results may motivate future work to characterize these α-SNAP variants, including their possible capacity to confer novel or enhanced resistance, and to help elucidate the molecular underpinnings of *rhg1-a* resistance.

A3.2 Results and Discussion

The α-SNAP$_{Rhg1}$ C-terminus is robust to many types of amino acid changes. The α-SNAP protein is composed of an N-terminal sheet and a C-terminal domain (Rice and Brunger, 1999). The C-terminal domain interacts with the AAA+ ATPase NSF, stimulating its activity to help disassemble SNARE bundles, allowing vesicle trafficking to proceed (Weidman et al., 1989; Barnard et al., 1996; May et al., 1999; Yu et al., 1999; Matveeva et al., 2002; Zhou et al., 2015a; Zhou et al., 2015b). Indeed, when the primary amino acid sequences for α-SNAP proteins are compared across species, the C-terminal domain is remarkably conserved; however, in α-SNAP$_{Rhg1}$ the sequence diverges (Bayless et al., 2016). It had previously been reported that the polymorphisms observed in α-SNAP$_{Rhg1}$ directly affect NSF interaction and subsequent vesicle trafficking (Bayless et al., 2016; Bayless et al., 2018). Using the methods described in those papers, we sought to learn which types of mutations were allowable and were able to disrupt

vesicle trafficking.

Using the α-SNAP$_{Rhg1}$ from a soybean accession carrying the commercially important PI88788 *Rhg1* haplotype as template, we sought to make 28 mutations in the C-terminal domain and test their effects in *N. benthamiana*. One model of *Rhg1*-mediated resistance suggests that disruption of vesicle trafficking, through impaired NSF/ α-SNAP interactions at nematode feeding sites, promotes feeding site collapse and subsequent resistance. The C-terminal domain of α-SNAP was therefore chosen as a target of mutation first as this is the interaction surface with the ATPase NSF. Additionally, the presence of nematode effectors encoding putative vesicle trafficking components (see below) suggests that modifying the α-SNAP may lead to reduced interaction with these nematode effectors with subsequently enhanced resistance. We made mutations with both similar (e.g., $Q_{285}N$) and dissimilar (e.g., $R_{230}E$) chemistries at these positions to explore the diversity of possible phenotypes as well as gain insight into the amino acid requirements of cell-death induction by α-SNAP$_{Rhg1}$. From this strategy, we were able to successfully generate 24 α-SNAP mutations that were then tested in *N. benthamiana*.

Remarkably, all 24 of the mutated α-SNAPs generated were able to cause cell death comparable to the α-SNAP found in PI88788 (Figure 1A and B). This suggests that the C-terminal region of α-SNAP$_{Rhg1}$ is fairly robust to point mutations without causing changes in protein stability that would cause significant degradation. This last point is further validated as immunoblots against the mutated α-SNAPs in a three gene context produced bands consistent with stably expressed protein (Figure 1C). These results provided impetus to test the α-SNAPs for their ability to confer resistance to otherwise virulent nematode populations (see below).

Because all tested α-SNAP$_{Rhg1}$HC with C-terminal point mutations were able to cause cell death in *N. benthamiana*, the data additionally suggest that the mechanism underlying cell death induced by α-SNAP$_{Rhg1}$HC (and probably α-SNAP$_{Rhg1}$LC) is not specific to the set of residues at the polymorphic sites. The small number of α-SNAP$_{Rhg1}$ variants that have been tested in previous studies interact equally poorly with the NSF in *N. benthamiana*, for example a previously tested LC$_{286-289}$AAAA mutant both interacted poorly with NSF$_{RAN07}$ and caused cell death in *N. benthamiana* (Bayless et al., 2018).

A likely hypothesis, in light of the large body of NSF/ α-SNAP structure/function data (i.e., Zhao et al. 2015a), is that these α-SNAPs with C-terminal region amino acid changes are able to bind SNARE bundles but fail to activate successful NSF-mediated SNARE bundle disassembly. Because SNARE/ α-SNAP/NSF 20S complexes are multimeric, one or a few mutant α-SNAPs per SNARE bundle may be sufficient to act in a dominant-negative fashion. We speculate that the mutant α-SNAPs tested in the present study bind SNARE bundles with sufficiently low dissociation constants that it becomes rare for SNARE bundles to be populated exclusively by wild-type α-SNAPs. Future work may be directed to elucidating the binding capacities of these α-SNAP variants with NSF$_{RAN07}$ using either binding assays with purified proteins, or cell-death suppression conferred by NSF$_{RAN07}$ in *N. benthamiana*. Such tests might further clarify the residues important for NSF$_{RAN07}$ interaction and guide generation of transgenic plants for SCN resistance assays.

Preliminary characterization of select mutations *rhg1-a* and *rhg1-b G. max* have distinct resistance profiles. While *rhg1-a* soybeans are able to resist HG type 2.5.7 nematode

populations, *rhg1-b* soybeans (with PI88788-type high-copy *Rhg1* loci), which dominate the currently available soybean cultivars available to U.S. farmers, are susceptible to HG type 2.5.7 nematode populations (Tylka, 2016). Though the mechanisms underlying this resistance/susceptibility difference between the two haplotypes are not well understood, the α-SNAPs are an attractive target both because they are different between the haplotypes and because SNARE-like proteins are a large class of SCN effectors (Cook et al., 2012; Cook et al., 2014; Lee et al., 2015; Bayless et al., 2016; Bayless et al., 2018; Gardner et al., 2018). We therefore sought to characterize the effects of a couple of the newly generated α-SNAP variants, L_{283}F and L_{283}A, by testing their ability to confer additional or novel resistance to HG type 2.5.7 nematodes in detached root assays.

From our preliminary data, the average number of nematodes on roots from both of the constructs was reduced relative to the control plants (Figure 2A). Moreover, when the number of nematodes of each developmental stage were compared between treatments, the percentage of nematodes past J2, typically a measure of overall resistance or susceptibility of a plant, was comparable between treatments (Figure 3B). Further, any resistance conferred by the additional or novel α-SNAP expression acted fairly uniformly at different developmental stages (Figure 2C). One limitation of the experiment is there was no control where additional α-$SNAP_{Rhg1}$HC was added to Fayette roots. Therefore, it is possible that additional expression of α-SNAP, and not the mutation, reduces the nematode numbers. Additional work should address whether it is additional expression of the α-SNAP that drives reduced nematode numbers. Additionally, the constructs used in future detached root assays should be guided by pretesting the construct for NSF_{RAN07} interaction.

From the data generated in *N. benthamiana*, it is clear that there are no obvious differences between α-SNAP$_{Rhg1}$HC and mutant variants in their ability to cause cell death though impairing vesicle trafficking. However, that there are no obvious effects resultant from expressing α-SNAP$_{Rhg1}$HC or mutant variants suggests that in soybean, the α-SNAP$_{Rhg1}$HC or mutants are likely not sufficient for complete resistance to HG type 2.5.7 nematode populations. This is consistent with prior observations that in *rhg1-a* soybean lines, which are HG type 2.5.7 resistant, the α-SNAP$_{Rhg1}$ only provides modest contributions to resistance against this nematode population (Meksem et al., 2001; Brucker et al., 2005; Wu et al., 2009; Yu et al., 2016). In these lines, it is SHMT$_{Rhg4}$ that is an additional contributor to HG type 2.5.7 resistance. Furthermore, there is some evidence suggesting that α-SNAP$_{Rhg1}$LC interacts with SHMT$_{Rhg4}$ to confer resistance (Kandoth et al., 207; Lakhssassi et al., 2020). Therefore, potential courses of research would be investigating the interactions between SHMT$_{Rhg4}$ and mutant α-SNAP variants, either looking for loss of resistance in transgenic *rhg1-a* roots expressing the variant or gain of resistance in *rhg1-b* roots co-expressing the SHMT$_{Rhg4}$ and α-SNAP variant.

A3.3 Methods

Vector Construction The open reading frame of soybean *Glyma.18g022500* (encoding α-SNAP$_{Rhg1}$) was obtained using cDNA generated from the cultivar Fayette and amplified through PCR using KapaHiFi polymerase (Kapa biosystems). PCR amplicons were purified using the Zymo Large Fragment DNA Recovery Kit (Zymo, Hercules CA) and assembled into pBlueScript containing the soybean ubiquitin promoter and nopaline synthase terminator through Gibson Assembly to generate pBlueScript-Fayette α-SNAP. The promoter-α-SNAP-*nosT* cassette was

then digested with PstI/XbaI (New England Biosciences). The digested fragment was purified using the Zymo Large Fragment DNA Recovery Kit (Zymo, Hercules CA), and ligated into the binary vector pSM101 using T4 DNA ligase (New England Biosciences). Mutations in the Fayette α-SNAP were generated similarly by the Kapa HiFi system with the pBlueScript Fayette α-SNAP plasmid as a template and digested with PstI/XbaI to generate the binary vector by subsequent ligation of the fragment into pSM101. For three-gene constructs, the mutated α-SNAP was cloned downstream of a CaMV 35S promoter in pSH190 (Figure 3) using ligase independent cloning (Huan et al., 1992).

Transient expression in Nb *Nicotiana benthamiana* plants were used for agro-infiltration assays. Constructs derived from pSM101 were used to transform *Agrobacterium tumefaciens* strain GV3101 (pMP90) using the freeze-thaw method. Transformed cells were cultured in LB with both kanamycin and rifampicin at concentrations of 25 μg/mL and 100 μg/mL, respectively. Cultured cells were harvested by centrifugation and resuspended in MMA induction buffer [1 L MMA: 5 g MS salts, 1.95 g MES, 20 g sucrose, 200 mM acetosyringone, pH = 5.6]. Suspensions were infiltrated into *Nb* using a tuberculin syringe with no needle. For cell death assays bacteria at OD_{600} = 0.6 were infiltrated into leaves and phenotypes were assessed 10 days after infiltration on a 0-5 scale: 0 (no cell death), 1 (mild and not uniform chlorosis throughout infiltrated area), 2 (chlorosis and some necrosis throughout infiltrated area), 3 (many flecks of necrosis throughout infiltrated area with some chlorosis), 4 (most of infiltrated area is necrotic), 5 (entire infiltrated area is necrotic).

Immunoblots and Antibodies Polyclonal rabbit antibodies raised against α-$SNAP_{Rhg1}$LC and α-$SNAP_{Rhg1}$HC were previously generated and validated as described in Bayless et al., 2016. Tissue preparation and immunoblots were performed essentially as in Bayless et al., 2016 and Bayless et al., 2018. In brief, *N. benthamiana* leaf tissue was flash frozen in liquid N_2, the mass measured, then homogenized using a PowerLyzer24 (Qiagen) for three to four cycles of 10 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% (vol/vol) Triton X-100, 10% (vol/vol) glycerol, 1/100 Sigma protease inhibitor cocktail for plants (P51200-1) was then added in a 3:1 volume to mass ratio. Lysates were then centrifuged at 7000 g and 7500 g for 10 minutes. Bradford assays were then performed on each supernatant and an equal amount of OD_{595} total protein was loaded in each sample lane of an SDS/PAGE gel. Immunoblots for α-SNAP and NSF 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 1:1000. The membrane was then washed with three washes in TBS-T. Secondary horseradish peroxidase conjugated goat anti-rabbit IgG (Sigma, A6145) was added at 1:10,000 in 5% non-fat dry milk in TBS-T and incubated for 1 hour at room temperature on a platform shaker followed by five washes with TBS-T. Chemiluminescence signal detection was performed with SuperSignal West Dura chemiluminescent substrate (Thermo Scientific) and developed using a Chemi Doc MP chemiluminescent imager (Bio-Rad).

Generation of transgenic soybean roots Soybean transgenic roots were generated essentially as in Cook et al., 2012, but using spectinomycin tolerance rather than GFP to screen for transgenic roots. Briefly *A. rhizogenes* strain Arqua1 was transformed with the pSH190

constructs using the freeze-thaw method and incubated at 28°C for two to four days. Fayette and Forrest soybean seeds were surface sterilized in a desiccator jar with chlorine gas for 16-20 hours. Sterilized seeds were then transferred to germination media (3.1 g/L Gamborg's B5 salts, 2% sucrose, l X Gaborg's B5 vitamins, 7% (w/v) agar, pH = 5.8) and germinated at 26°C (18hr light/6hr dark) for between five and seven days, then cotyledons harvested from the plants. Cotyledons were transformed by making several shallow slices on the abaxial side of the cotyledon, using a scalpel that was dipped in *A. rhizogenes* solution (OD_{600} between 0.55 and 0.7) Transformed cotyledons were then placed abaxial side down on filter paper on co-culture media (0.31 g/L Gamborg's B5 salts, 3% sucrose, 1 X Gamborg's B5 vitamins, 0.4 g/L L-cysteine, 0.154 g/L DTT, 0.245 g/L sodium thiosulfate, 40 mg/L acetosyringone, 5% agar, pH = 5.4), and incubated in the dark at room temperature for three days. Cotyledons were then transferred adaxial side down onto hairy root media with spectinomycin (4.3 g/L MS salts, 2% sucrose, 1 X Gamborg's B5 vitamins, 7% agar, 0.15 g/L cefotaxamine, 0.15 g/L carbenicillin, 150 mg/L spectinomycin, pH = 5.6), and incubated in the dark at room temperature for approximately two to three weeks. Transformed roots were selected based on robust elongation in the presence of spectinomycin.

Nematode demographic assays Demographic assays were essentially carried out as in Melito et al., 2010 and Cook et al., 2012. In brief, approximately 200-250 HG type 2.5.7 J2 nematodes in 0.05% agarose were inoculated onto transformed roots after sufficient elongation. Nematode infection was monitored using acid fuchsin staining, approximately two to three weeks after inoculation. Here, root segments were removed from agar and bleached in a 10% bleach

solution for 15 minutes. Bleached roots were rinsed twice in DI water before being placed in a 1% acetic acid solution for 15 minutes. Roots were then placed into an acid fuchsin staining solution (24 mL stock solution (0.35 g acid fuchsin, 25 mL glacial acetic acid, 75 mL ddH_20), 976 mL ddH_20), preheated to 100°C for approximately two minutes. Roots were then destained in ddH_20 for 30 minutes before being mounted in acidified glycerol (1 L: 2 mL HCl, 700 mL glycerol, 298 mL ddH_20). Numbers of nematodes that were at J2, J3 or a more advanced growth stage was then estimated based on morphology.

Primers Used in Appendix

H286Y FW: TACGAGGCTATTACTTGAAAAGGGCG

H286Y RV: AGTAATAGCCTCGTACTGCTCAAGTTCTTT

H286K FW: AAGGAGGCTATTACTTGAAAAGGGCG

H286K RV: AGTAATAGCCTCCTTCTGCTCAAGTTCTTT

H286F FW: TTCGAGGCTATTACTTGAAAAGGGCG

H286F RV: AGTAATAGCCTCGAACTGCTCAAGTTCTTT

H286W FW: TGGGAGGCTATTACTTGAAAAGGGCG

H286W RV: AGTAATAGCCTCCCACTGCTCAAGTTCTTT

Q285N FW: AACCATGAGGCTATTACTTGAAAAGG

Q285N RV: AATAGCCTCATGGTTCTCAAGTTCTTTGGC

Q285F FW: TTCCATGAGGCTATTACTTGAAAAGG

Q285F RV: AATAGCCTCATGGAACTCAAGTTCTTTGGC

Q285E FW: GAACATGAGGCTATTACTTGAAAAGG

Q285E RV: AATAGCCTCATGTTCCTCAAGTTCTTTGGC

L283I FW: ATCGAGCAGCATGAGGCTATTACTTG

L283I RV: CTCATGCTGCTCGATTTCTTTGGCTTTCAG

L283A FW: GCTGAGCAGCATGAGGCTATTACTTG

L283A RV: CTCATGCTGCTCAGCTTCTTTGGCTTTCAG

L283F FW: TTCGAGCAGCATGAGGCTATTACTTG

L283F RV: CTCATGCTGCTCGAATTCTTTGGCTTTCAG

E282D FW: GATCTTGAGCAGCATGAGGCTATTAC

E282D RV: ATGCTGCTCAAGATCTTTGGCTTTCAGCTT

E282F FW: TTCCTTGAGCAGCATGAGGCTATTAC

E282F RV: ATGCTGCTCAAGGAATTTGGCTTTCAGCTT

K281R FW: CGAGAACTTGAGCAGCATGAGGCTAT

K281R RV: CTGCTCAAGTTCTCGGGCTTTCAGCTTTTC

K281H FW: CATGAACTTGAGCAGCATGAGGCTAT

K281H RV: CTGCTCAAGTTCATGGGCTTTCAGCTTTTC

K281F FW: TTCGAACTTGAGCAGCATGAGGCTAT

K281F RV: CTGCTCAAGTTCGAAGGCTTTCAGCTTTTC

A280L FW: TTAAAAGAACTTGAGCAGCATGAGGC

A280L RV: CTCAAGTTCTTTTAATTTCAGCTTTTCCTT

A280V FW: GTTAAAGAACTTGAGCAGCATGAGGC

A280V RV: CTCAAGTTCTTTAACTTTCAGCTTTTCCTT

A280F FW: TTCAAAGAACTTGAGCAGCATGAGGC

A280F RV: CTCAAGTTCTTTGAATTTCAGCTTTTCCTT

K279R FW: CGAGCCAAAGAACTTGAGCAGCATGA

K279R RV: AAGTTCTTTGGCTCGCAGCTTTTCCTTCAC

K279H FW: CATGCCAAAGAACTTGAGCAGCATGA

K279H RV: AAGTTCTTTGGCATGCAGCTTTTCCTTCAC

K279F FW: TTCGCCAAAGAACTTGAGCAGCATGA

K279F RV: AAGTTCTTTGGCGAACAGCTTTTCCTTCAC

K279E FW: GAAGCCAAAGAACTTGAGCAGCATGA

K279E RV: AAGTTCTTTGGCTTCCAGCTTTTCCTTCAC

K203R FW: CGACTCTGTAAAGAGGACGTTGTTGC

K203R RV: CTCTTTACAGAGTCGGCAGATGCCAGCATT

K203E FW: GAACTCTGTAAAGAGGACGTTGTTGC

K203E RV: CTCTTTACAGAGTTCGCAGATGCCAGCATT

H286P FW: CCAGAGGCTATTACTTGAAAAGGGCG

H286P RV: AGTAATAGCCTCTGGCTGCTCAAGTTCTTT

Q285D FW: GACCATGAGGCTATTACTTGAAAAGG

Q285D RV: AATAGCCTCATGGTCCTCAAGTTCTTTGGC

R230E FW: GAAGAATATAGATTGTTGGCGGACAT

R230E RV: CAATCTATATTCTTCTGTTCCTGAAAATGT

K193E FW: GAAGGACACCTTCTTAATGCTGGCAT

K193E RV: AAGAAGGTGTCCTTCAACTCCATACTTCAG

H286Y LIC: GAGGAGAAGAGCCGTCAAGTAATAGCCTCGTACTGCTCA

H286K LIC: GAGGAGAAGAGCCGTCAAGTAATAGCCTCCTTCTGCTCA

H286F LIC: GAGGAGAAGAGCCGTCAAGTAATAGCCTCGAACTGCTCA

H286W LIC: GAGGAGAAGAGCCGTCAAGTAATAGCCTCCCACTGCTCA

H286P LIC: GAGGAGAAGAGCCGTCAAGTAATAGCCTCTGGCTGCTCA

Q285N LIC: GAGGAGAAGAGCCGTCAAGTAATAGCCTCATGGTTCTCA

Q285F LIC: GAGGAGAAGAGCCGTCAAGTAATAGCCTCATGGAACTCA

Q285E LIC: GAGGAGAAGAGCCGTCAAGTAATAGCCTCATGTTCCTCA

L283I LIC: GAGGAGAAGAGCCGTCAAGTAATAGCCTCATGCTGCTCG

L283A LIC: GAGGAGAAGAGCCGTCAAGTAATAGCCTCATGCTGCTC

WT LIC: GAGGAGAAGAGCCGTCAAGTAATAGCCTCATGCTGCTCA

A3.4 References

Barnard, R.J., Morgan, A. and Burgoyne, R.D. (1996). Domains of alpha-SNAP required for the stimulation of exocytosis and for N-ethylmalemide-sensitive fusion protein (NSF) binding and activation. *Molecular Biology of the Cell*, 7(5), pp.693–701.

Bayless, A.M., Smith, J.M., Song, J., McMinn, P.H., Teillet, A., August, B.K. and Bent, A.F. (2016). Disease resistance through impairment of α-SNAP–NSF interaction and vesicular trafficking by soybean Rhg1. *Proceedings of the National Academy of Sciences*, 113(47), pp.E7375–E7382.

Bayless, A.M., Zapotocny, R.W., Grunwald, D.J., Amundson, K.K., Diers, B.W. and Bent, A.F. (2018). An atypical N-ethylmaleimide sensitive factor enables the viability of nematode-resistant Rhg1 soybeans. *Proceedings of the National Academy of Sciences*, 115(19), pp.E4512–E4521.

Bekal, S., Domier, L.L., Gonfa, B., Lakhssassi, N., Meksem, K. and Lambert, K.N. (2015). A SNARE-Like Protein and Biotin Are Implicated in Soybean Cyst Nematode Virulence. *PLOS ONE*, 10(12), p.e0145601.

Brucker, E., Carlson, S., Wright, E., Niblack, T. and Diers, B. (2005). Rhg1 alleles from soybean PI 437654 and PI 88788 respond differentially to isolates of Heterodera glycines in the greenhouse. *Theoretical and Applied Genetics*, 111(1), pp.44–49.

Cook, D.E., Bayless, A.M., Wang, K., Guo, X., Song, Q., Jiang, J. and Bent, A.F. (2014). Distinct Copy Number, Coding Sequence, and Locus Methylation Patterns Underlie Rhg1-Mediated Soybean Resistance to Soybean Cyst Nematode. *Plant Physiology*, 165(2), pp.630–647.

Cook, D.E., Lee, T.G., Guo, X., Melito, S., Wang, K., Bayless, A.M., Wang, J., Hughes, T.J., Willis, D.K., Clemente, T.E., Diers, B.W., Jiang, J., Hudson, M.E. and Bent, A.F. (2012). Copy Number Variation of Multiple Genes at Rhg1 Mediates Nematode Resistance in Soybean. *Science*, 338(6111), pp.1206–1209.

Dong, J., Zielinski, R.E. and Hudson, M.E. (2020). t-SNAREs Bind the Rhg1 α-SNAP and Mediate Soybean Cyst Nematode Resistance. *The Plant Journal*.

Ford, K., McDonald, D. and Mali, P. (2019). Functional Genomics via CRISPR–Cas. *Journal of Molecular Biology*, 431(1), pp.48–65.

Gardner, M., Dhroso, A., Johnson, N., Davis, E.L., Baum, T.J., Korkin, D. and Mitchum, M.G. (2018). Novel global effector mining from the transcriptome of early life stages of the soybean cyst nematode Heterodera glycines. *Scientific Reports*, 8(1).

Huan, R., Serventi, I. and Moss, J. (1992). Rapid, reliable ligation-independent cloning of PCR products using modified plasmid vectors. *Biotechniques*, 13(4), pp.515–518.

Kandoth, P.K., Liu, S., Prenger, E., Ludwig, A., Lakhssassi, N., Heinz, R., Zhou, Z., Howland, A., Gunther, J., Eidson, S., Dhroso, A., LaFayette, P., Tucker, D., Johnson, S., Anderson, J., Alaswad, A., Cianzio, S.R., Parrott, W.A., Korkin, D., Meksem, K. and Mitchum, M.G. (2017). Systematic Mutagenesis of Serine Hydroxymethyltransferase Reveals an Essential Role in Nematode Resistance. *Plant Physiology*, 175(3), pp.1370–1380.

Lakhssassi, N., Piya, S., Bekal, S., Liu, S., Zhou, Z., Bergounioux, C., Miao, L., Meksem, J., Lakhssassi, A., Jones, K., Kassem, M.A., Benhamed, M., Bendahmane, A., Lambert, K., Boualem, A., Hewezi, T. and Meksem, K. (2020). A pathogenesis-related protein GmPR08-Bet VI promotes a molecular interaction between the GmSHMT08 and GmSNAP18 in resistance to Heterodera glycines. *Plant Biotechnology Journal*, 18(8), pp.1810–1829.

Lee, T.G., Kumar, I., Diers, B.W. and Hudson, M.E. (2015). Evolution and selection ofRhg1,a copy-number variant nematode-resistance locus. *Molecular Ecology*, 24(8), pp.1774–1791.

Lobingier, B.T., Nickerson, D.P., Lo, S.-Y. and Merz, A.J. (2014). SM proteins Sly1 and Vps33 co-assemble with Sec17 and SNARE complexes to oppose SNARE disassembly by Sec18. *eLife*, 3.

Marz, K.E., Lauer, J.M. and Hanson, P.I. (2003). Defining the SNARE Complex Binding Surface of α-SNAP. *Journal of Biological Chemistry*, 278(29), pp.27000–27008.

Matveeva, E.A., May, A.P., He, P. and Whiteheart, S.W. (2002). Uncoupling the ATPase Activity of the N-Ethylmaleimide Sensitive Factor (NSF) from 20S Complex Disassembly†. *Biochemistry*, 41(2), pp.530–536.

May, A.P., Misura, K.M.S., Whiteheart, S.W. and Weis, W.I. (1999). Crystal structure of the amino-terminal domain of N-ethylmaleimide-sensitive fusion protein. *Nature Cell Biology*, 1(3), pp.175–182.

McCallum, C.M., Comai, L., Greene, E.A. and Henikoff, S. (2000). Targeting Induced Local Lesions IN Genomes (TILLING) for Plant Functional Genomics. *Plant Physiology*, [online] 123(2), pp.439–442. Available at: http://www.plantphysiol.org/content/123/2/439 [Accessed 30 Apr. 2019].

Meksem, K., Pantazopoulos, P., Njiti, V.N., Hyten, L.D., Arelli, P.R. and Lightfoot, D.A. (2001). 'Forrest' resistance to the soybean cyst nematode is bigenic: saturation mapping of the Rhg1and Rhg4 loci. *Theoretical and Applied Genetics*, 103(5), pp.710–717.

Melito, S., Heuberger, A.L., Cook, D., Diers, B.W., MacGuidwin, A.E. and Bent, A.F. (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(1), p.104.

Morey, C., Kienle, C.N., Klöpper, T.H., Burkhardt, P. and Fasshauer, D. (2017). Evidence for a conserved inhibitory binding mode between the membrane fusion assembly factors Munc18 and syntaxin in animals. *Journal of Biological Chemistry*, 292(50), pp.20449–20460.

Muteeb, G. and Sen, R. (2010). Random Mutagenesis Using a Mutator Strain. *Methods in Molecular Biology*, pp.411–419.

Rice, L.M. and Brunger, A.T. (1999). Crystal Structure of the Vesicular Transport Protein Sec17. *Molecular Cell*, 4(1), pp.85–95.

Song, H., Orr, A., Duan, M., Merz, A.J. and Wickner, W. (2017). Sec17/Sec18 act twice, enhancing membrane fusion and then disassembling cis-SNARE complexes. *eLife*, 6.

Stepien, K.P., Prinslow, E.A. and Rizo, J. (2019). Munc18-1 is crucial to overcome the inhibition of synaptic vesicle fusion by αSNAP. *Nature Communications*, 10(1).

Tylka, G.L. (2016). Understanding Soybean Cyst Nematode HG Types and Races. *Plant Health Progress*, 17(2), pp.149–151.

Weidman, P.J., Melançon, P., Block, M.R. and Rothman, J.E. (1989). Binding of an N-ethylmaleimide-sensitive fusion protein to Golgi membranes requires both a soluble protein(s) and an integral membrane receptor. *The Journal of Cell Biology*, 108(5), pp.1589–1596.

Wu, X., Blake, S., Sleper, D.A., Shannon, J.G., Cregan, P. and Nguyen, H.T. (2009). QTL, additive and epistatic effects for SCN resistance in PI 437654. *Theoretical and Applied Genetics*, 118(6), pp.1093–1105.

Yu, N., Lee, T.G., Rosa, D.P., Hudson, M. and Diers, B.W. (2016). Impact of Rhg1 copy number, type, and interaction with Rhg4 on resistance to Heterodera glycines in soybean. *Theoretical and Applied Genetics*, 129(12), pp.2403–2412.

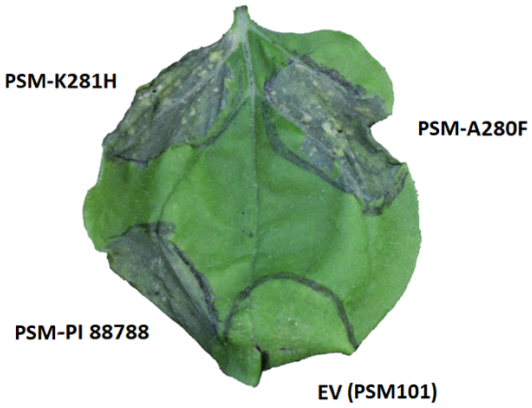
Yu, R.C., Jahn, R. and Brunger, A.T. (1999). NSF N-Terminal Domain Crystal Structure. *Molecular Cell*, 4(1), pp.97–107.

Zhao, M., Wu, S., Zhou, Q., Vivona, S., Cipriano, D.J., Cheng, Y. and Brunger, A.T. (2015a). Mechanistic insights into the recycling machine of the SNARE complex. *Nature*, 518(7537), pp.61–67

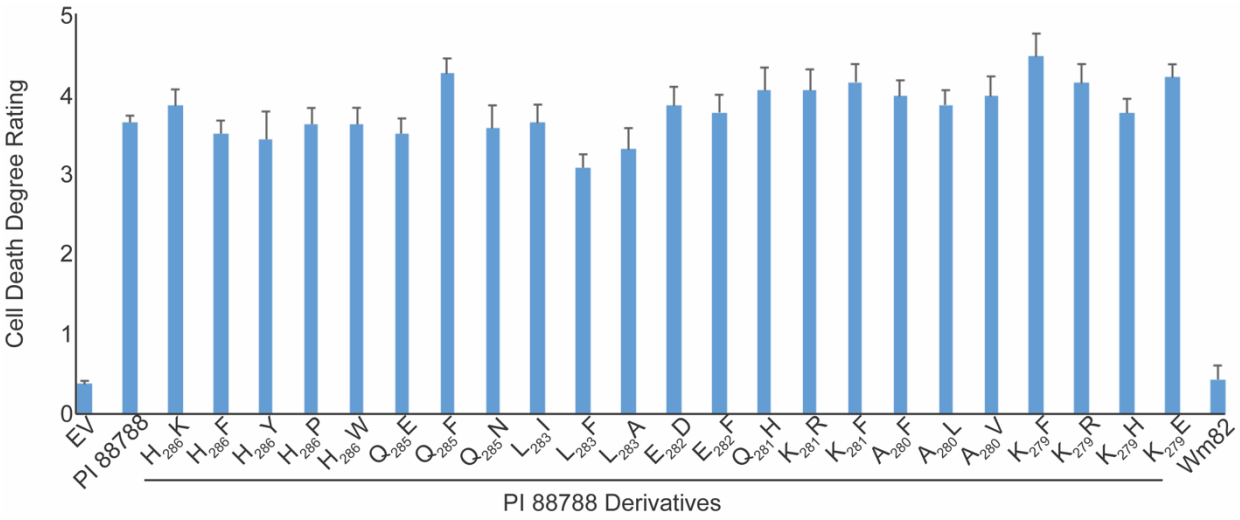
Zhou, Q., Huang, X., Sun, S., Li, X., Wang, H.-W. and Sui, S.-F. (2015b). Cryo-EM structure of SNAP-SNARE assembly in 20S particle. *Cell Research*, 25(5), pp.551–560.

A3.5 Figures

A



B



C

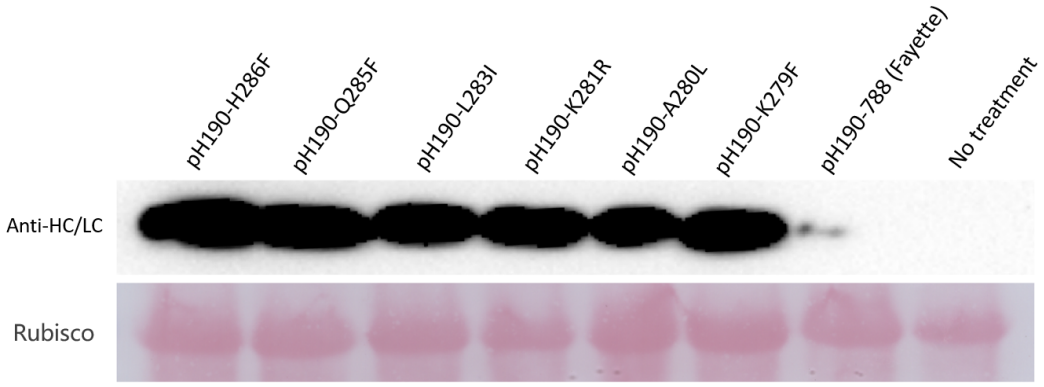
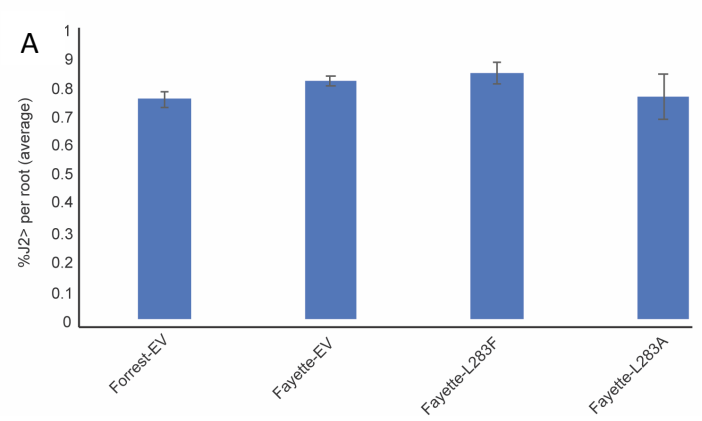
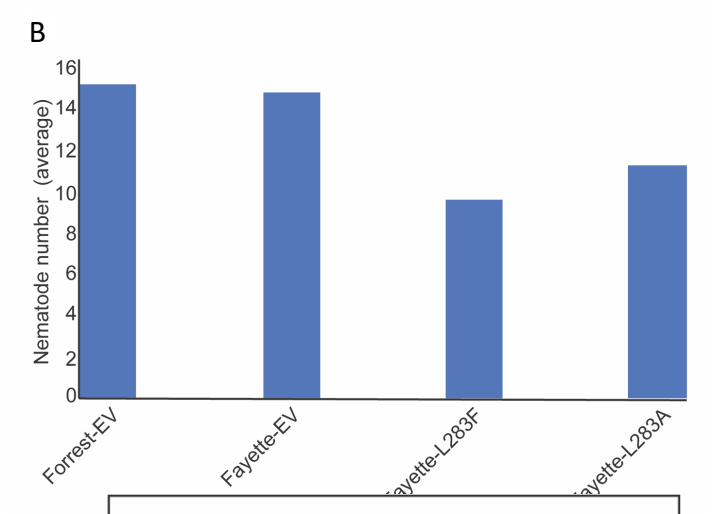


Figure 1: α-SNAP$_{Rhg1}$HC variants are robust to C-terminal amino acid mutations (A) representative leaf of *N. benthamiana* infiltrated with *A. tumefaciens* GV3101 containing the denoted α-SNAP variant in pSM101 after seven days, showing necrosis throughout areas infiltrated with α-SNAP$_{Rhg1}$HC variants. (B) Summary of cell death ratings of 24 of the tested α-SNAP$_{Rhg1}$HC variants. (C) Immunoblot of leaf tissue from *N. benthamiana* infiltrated with *A. tumefaciens* GV3101 containing the denoted α-SNAP variant in pSH190. There is no gene encoding an α-SNAP in the PI 88788/Fayette plasmid.


A
%J2> per root (average)
1
9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Forrest-EV
Fayette-EV
Fayette-L283F
Fayette-L283A


B
Nematode number (average)
16
14
12
10
8
6
4
2
0
Forrest-EV
Fayette-EV

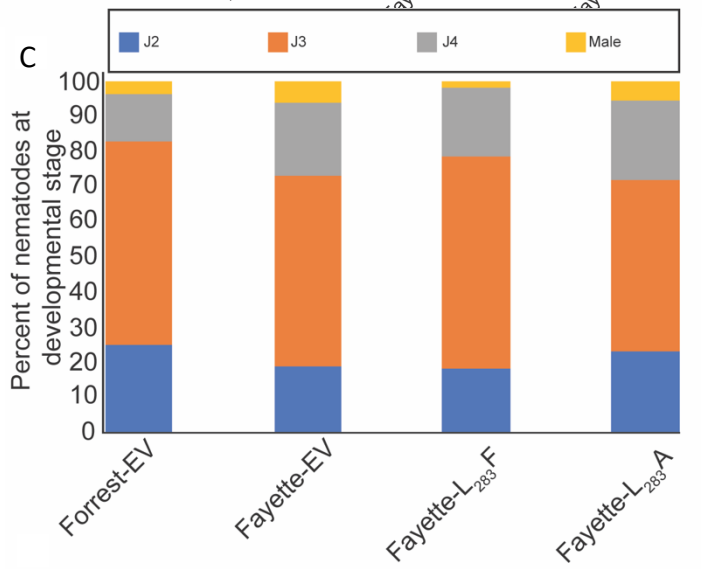

C
J2
J3
J4
Male
Percent of nematodes at developmental stage
100
90
80
70
60
50
40
30
20
10
0
Forrest-EV
Fayette-EV
Fayette-L283F
Fayette-L283A

Figure 2: α-SNAP$_{Rhg1}$HC variant expression reduces average nematode numbers but not development. (A) average number of HG type 2.5.7 nematodes (J2-J4 and males) in detached roots of denoted genotype expressing designated α-SNAP in pSH190 or empty vector. (B) percentage of nematode development past J2 from data collected in A. (C) percentage of nematodes at each developmental stage and males from data collected in A. Blue=J2, Orange=J3, Gray=J4, Yellow=male

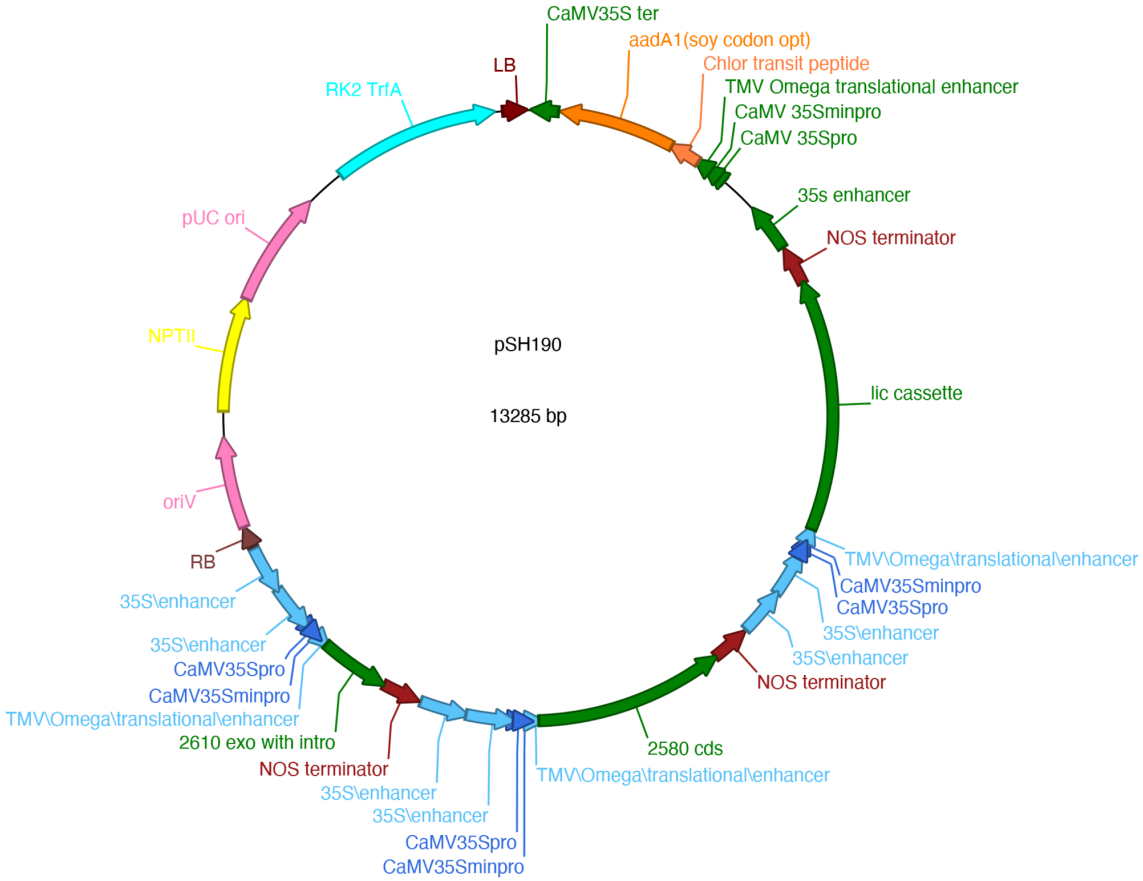


Figure 3: pSH190 binary vector map pSH190 is a binary vector containing a 2580, 2610 and a LIC cassette for introduction of the α-SNAP variants, each driven by CaMV 35S promoters. Bacterial selection is kanamycin, and plant selection is spectinomycin.

Appendix 4: Leveraging yeast to study effector interactions with an example from *Ralstonia solanacearum*

Derrick Grunwald, Connor Hendrich, Caitilyn Allen, Andrew Bent

For HG type 0 nematode effectors, I performed all cloning and preliminary assays. For cloning and characterization of *Ralstonia solanacearum* RipAX1 and RipAX2 effectors, I was assisted in cloning and characterization by Connor Hendrich.

A4.1 Introduction

Many bacterial pathogens produce effector molecules which target essential eukaryotic biochemical processes including central metabolism, cytoskeletal dynamics, vesicle trafficking, signal transduction and replication (Dolinski and Botstein, 2007; Siggers and Lesser, 2008; Curak et al., 2009; Botstein and Fink, 2011). As these processes are conserved across eukaryotes, effector molecules which interact with essential processes in a natural host may also interact with the same process in yeast (*Saccharomyces cerevisiae*) systems. This is especially attractive as yeast are amenable to genetic manipulation as well as the observation that many bacterial proteins do not affect yeast growth and replication, meaning that any impacts are due to the

expressed protein and not from heterologous expression (Slagowski et al., 2008). Such approaches have yielded insights into the mechanisms underlying pathogenesis from organisms as varied as *Pseudomonas syringae*, *Vibrio cholerae*, *Legionella*, and *R. solanacearum* (Munkvold et al., 2008; O'Brien et al., 2015; Sreelatha et al., 2015; Popa et al., 2016).

In addition to heterologous expression, expression of effector molecules in yeast may impact growth and reproduction only under stress conditions. Commonly targeted pathways include the cell wall integrity (though addition of caffeine), high osmolarity/glycerol (through addition of sorbitol or ionic salts), the unfolded protein response (through addition of tunicamycin), and target of rapamycin (through addition of caffeine) (Gustin et al., 1998; Hohmann, 2002; Parsons et al., 2004; Kuranda et al., 2006; Auesukaree et al., 2009). Indeed, integration of chemical genetic approaches have yielded insights into the functions of *P. syringae* and *Xanthomonas campestis* pv. *vesicatoria* effectors which otherwise produced no phenotype (Salomon et al., 2010; Salomon et al., 2012a; Salomon et al., 2012b; Lee et al., 2019).

R. solanacearum is a soil-borne pathogen that is the causal agent of bacterial wilt on important crop species such as potato and tomato, although it has many additional plant hosts (Salanoubat et al., 2002). Due to its wide geographic distribution and many potential hosts, *R. solanacearum* has been described as the second most important bacterial pathogen (Mansfield et al., 2012). However, in spite of its agronomic importance, relatively little is known about the effector molecules that underly pathogenesis in this organism (Landry et al., 2020). In fact, of the more than 70 type III effectors predicted to be in *R. solanacearum*, only a handful have been characterized (Nahar et al., 2013; Peeters et al., 2013; Popa et al., 2016; Landry et al.,

2020). Of these, RipAX2/Rip36 is a putative zinc metalloprotease whose recognition is implicated in nonhost resistance to *R. solanacearum*, though the role of the effector in host plants remains to be elucidated (Nahar et al., 2013; Morel et al., 2018).

Below, we briefly describe heterologous expression of *R. solanacearum* – encoded RipAX2/Rip36 and the RipAX1 paralog in the yeast model system *S. cerevisiae* with the intention of beginning to elucidate potentially affected pathways. Furthermore, we use the yeast system to investigate the ability of HG type 0 nematode effectors to induce yeast cell death. In both cases, we test effector expression in the presence or absence of select stressors to determine if a particular yeast pathway is targeted by said effector.

A4.2 Results and discussion

Many plant pathogen effectors target processes conserved throughout the eukaryotic domain of life. Buoyed by recent success in determining targets of effectors in yeast systems for *P. syringae* and *R. solanacearum*, we sought to apply the same methodologies for effectors derived from an HG type 0 nematode population. Additionally, we added stressors of the yeast unfolded protein response (tunicamycin), high osmolarity/glycerol pathway (NaCl and sorbitol), and cell wall integrity and target of rapamycin pathways (caffeine) to determine if any of the effectors could be synthetically lethal with the stressor, indicating a pathway potentially targeted by the effector. We first tested eight SCN effectors along with BAX (positive control) and an empty vector (negative control) for impaired cell growth when expression was induced in the yeast strain RSY255. BAX was able to impair yeast cell growth when induced, and this phenotype was exacerbated when expressed in the presence of a stressor (Figure 1). Empty

vector did not impair yeast growth when induced alone or in the presence of any stressors. Of the HG type 0 nematode effectors screened, none were able to impair yeast cell growth when induced alone or in the presence of stressor. This included 2A05 and 2B10, which garnered attention in *N. benthamiana* α-SNAP$_{Rhg1}$ cell death suppression assays (Appendix 1).

That none of the tested effectors were able to induce cell death is perhaps not surprising as a recent publication suggests that many HG type 0 nematode effectors are able to suppress PCD pathways in yeast (Wang et al., 2020). This PCD was induced by BAG6, which had previously been shown to be highly induced in the syncytia of resistant *G. max* (Kandoth et al., 2011). In Figure 2, we show that the α-SNAP$_{Rhg1}$ is able to impair yeast cell growth when induced alone or in the presence of stressors. Future work should seek to determine whether the effectors that others have reported able to suppress BAG6-induced PCD are similarly able to suppress α-SNAP$_{Rhg1}$-induced suppression of yeast growth. Previously untested SCN effectors can also be studied using the above regime.

Along with HG type 0 nematode effectors, we tested the ability of the *R. solanacearum* effectors RipAX1 and RipAX2 to induce changes in yeast growth in the presence or absence of stresses. Interestingly, RipAX1 was able to attenuate growth of yeast when in the presence of sorbitol, but not other stressors (Figure 2). This suggests that RipAX1 might target the high-osmolarity glycerol pathway. Consistent with this observation is the finding that high-osmolarity glycerol signaling was similarly attenuated in yeast by the *P. syringae* effector HopX1 (Salomon et al., 2012). Future work should be directed to elucidating the mechanism underlying this attenuated growth. Toward these ends, testing expression of Hog1 through the Hog1-

responsive CRE promoter element-lacZ fusions or testing the localization dynamics of Hog1-GFP fusions in the presence of RipAX1 are attractive experiments (Proft et al., 2001).

For both HG type 0 and *R. solanacearum* effectors, it is possible that effectors act through altering vesicle trafficking dynamics. This is especially pertinent as SNARE-like proteins are among the largest and most variable class of nematode effectors, and the observation that RipAX1 and the BoNT (which cleaves mammalian SNARE bundles) from *Clostridium botulinum* both belong to the zinc metalloprotease protein family (Gardner et al., 2018). Direct visualization of yeast vacuole fragmentation and accumulation of ER would be suggestive of vesicle trafficking dysfunction. Additionally, synthetic lethality of an effector expressed in a temperature sensitive secretion mutant at permissive temperatures could indicate vesicle trafficking dysfunction. Conversely, the ability of an effector to rescue any of the temperature-sensitive secretion mutants at non-permissive temperatures, although unlikely, would suggest a restoration of vesicle trafficking. Such experiments may likewise be complemented *in planta* through sec-GFP experiments in a stable *N. benthamiana* or *G. max* line. Further work may additionally utilize the generated yeast two-hybrid or split-ubiquitin vectors to test for direct binary interactions between the effector of interest and components of either the vesicle trafficking or Hog pathway components.

A4.3 Methods

Vector Construction Genomic DNA from *R. solanacearum* was extracted using Epicentere bacterial gDNA extraction kit (Lucigen) using manufacturer's instructions. The *RipAX1* and *RipAX2* genes were amplified by PCR using the KapaHiFi polymerase (Kapa Biosystems). The

amplicons were then purified using the Zymo Large Fragment DNA Recovery Kit (Zymo, Hercules CA), and assembled into the galactose-inducible vector pSH221 (Figure 3) by ligase independent cloning (Huan et al., 1992). LIC amplicons for *BAX* (from pGilda-BAX; Kampranis et al., 2000), *GmSNAP18* and *2580* were generated similarly. Amplicons from effectors generated in Appendix 1 were similarly used. For split-ubiquitin and yeast two-hybrid vector construction, amplicons were cloned into pSH219, WY145 and WY146 by ligase independent cloning (Figure 4A-C)

Yeast Inhibition assays pSH221-based expression vectors carrying *RipAX1* or other genes were transformed into the *S. cerevisiae* strains EGY48 (MATa, *his3*, *trp1*, *ura3*, LexAop(x6)-LEU2; Frick and Johnston, 1990) or RSY255 (MATa, *ura3-52 leu2-3, -112*) by the lithium acetate method. In brief, 300 ng of vector, 100 uL TEL [100 mL: 80 mL ddH_20, 10 mL 10 X TE (100 mL: 10 mL 1 M Tris-HCl, 2 mL 0.5 M EDTA, 88 mL H_20, pH = 7.5), 10 mL 1 M LiOAc], 1 mL of PEG-TEL (100 mL: 10 mL 10X TE, 10 mL LiOAc, 80 mL 50% (w/v) PEG), and 4 uL of sheared salmon sperm DNA (CAS# AM9680, ThermoFischer) were incubated overnight at room temperature. The next day, the yeast were heat shocked for 10 minutes at 42°C, spun down and resuspended in 250 uL of sterile water. Transformed yeast were plated onto yeast minimal media [1 L: 20 g glucose, 20 g agar, 10 mL 100 X amino acid supplement (100 mL: 300 mg adenine, 200 mg histidine, 400 mg leucine, 400 mg lysine, 300 mg tryptophan, 200 mg uracil, 100 mL ddH_20), 990 mL ddH_20] lacking uracil and incubated at 30°C until colonies appeared. Growth inhibition assays were followed essentially as in Salomon et al. (2011). Briefly, colonies were purified and grown in liquid yeast minimal media lacking uracil at 30°C overnight. The yeast was then pelleted,

washed, and normalized to an OD_{600} of 1.0. Each strain was then 10-fold serially diluted four times and spotted (5 uL) onto repressing (yeast minimal media with 20 g glucose) or inducing (yeast minimal media with 20 g galactose and 10 g raffinose) media supplemented with caffeine (7 mM), sodium chloride (0.5 M), sorbitol (1 M), or tunicamycin (0.12 ug/mL) and incubated either at 30°C or 37°C. Phenotypes were visualized 5 days after plating.

Primers Used in Appendix

AX2_LIC_FW: CGACGACAAGACCGTGCACCATGCTTATACAGACACAGTACCCTG

AX2_LIC_REV: GAGGAGAAGAGCCGTTTTGCGTTGCGTGGCTTGTAC

AX1_LIC_FW: CGACGACAAGACCGTGCACCATGCCGATTGAAACCCGATACAGC

AX1_LIC_REV: GAGGAGAAGAGCCGTTTATAGTATTTGCGTCGCAAGGG

BAX_LIC_FW: GAGGAGAAGAGCCGTCAGCCCATCTTCTTCCAGATGGT

BAX_LIC_REV: ATTAACAAGGCCATTACGGCCTCCACTATTAATTCGGTAACCGTACAAGTGAAC

A4.4 References

Auesukaree, C., Damnernsawad, A., Kruatrachue, M., Pokethitiyook, P., Boonchird, C., Kaneko, Y. and Harashima, S. (2009). Genome-wide identification of genes involved in tolerance to various environmental stresses inSaccharomyces cerevisiae. *Journal of Applied Genetics*, 50(3), pp.301–310.

Botstein, D. and Fink, G.R. (2011). Yeast: An Experimental Organism for 21st Century Biology. *Genetics*, 189(3), pp.695–704.

Curak, J., Rohde, J. and Stagljar, I. (2009). Yeast as a tool to study bacterial effectors. *Current Opinion in Microbiology*, 12(1), pp.18–23.

Dolinski, K. and Botstein, D. (2007). Orthology and Functional Conservation in Eukaryotes. *Annual Review of Genetics*, 41(1), pp.465–507.

Flick, J.S. and Johnston, M. (1990). Two systems of glucose repression of the GAL1 promoter in Saccharomyces cerevisiae. *Molecular and Cellular Biology*, 10(9), pp.4757–4769.

Gardner, M., Dhroso, A., Johnson, N., Davis, E.L., Baum, T.J., Korkin, D. and Mitchum, M.G. (2018). Novel global effector mining from the transcriptome of early life stages of the soybean cyst nematode Heterodera glycines. *Scientific Reports*, 8(1).

Gustin, M., Albertyn, J., Alexander, M. and Davenport, K. (1998). MAP Kinase Pathways in the Yeast Saccharomyces cerevisiae. *Microbiology and Molecular Biology Reviews*, 62(4), pp.1264–1300.

Hohmann, S. (2002). Osmotic Stress Signaling and Osmoadaptation in Yeasts. *Microbiology and Molecular Biology Reviews*, 66(2), pp.300–372.

Huan, R., Serventi, I. and Moss, J. (1992). Rapid, reliable ligation-independent cloning of PCR products using modified plasmid vectors. *Biotechniques*, 13(4), pp.515–518.

Kampranis, S., Damianova, R., Atallah, M., Toby, G., Kondi, G., Tsichlis, P. and Markis, A. (2000). A Novel Plant Glutathione S-Transferase/Peroxidase Suppresses Bax Lethality in Yeast. *Journal of Biological Chemistry*, 275(38), pp.29207–29216.

Kandoth, P.K., Ithal, N., Recknor, J., Maier, T., Nettleton, D., Baum, T.J. and Mitchum, M.G. (2011). The Soybean Rhg1 Locus for Resistance to the Soybean Cyst Nematode Heterodera glycines Regulates the Expression of a Large Number of Stress- and Defense-Related Genes in Degenerating Feeding Cells. *Plant Physiology*, 155(4), pp.1960–1975.

Kuranda, K., Leberre, V., Sokol, S., Palamarczyk, G. and Francois, J. (2006). Investigating the caffeine effects in the yeast Saccharomyces cerevisiae brings new insights into the connection between TOR, PKC and Ras/cAMP signaling pathways. *Molecular Microbiology*, [online] 61(5), pp.1147–1166. Available at: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2958.2006.05300.x [Accessed 2 Dec. 2019].

Landry, D., González-Fuente, M., Deslandes, L. and Peeters, N. (2020). The large, diverse, and robust arsenal of Ralstonia solanacearum type III effectors and their in planta functions. *Molecular Plant Pathology*, 21(10), pp.1377–1388

Lee, A.H.-Y., Bastedo, D.P., Youn, J.-Y., Lo, T., Middleton, M.A., Kireeva, I., Lee, J.Y., Sharifpoor, S., Baryshnikova, A., Zhang, J., Wang, P.W., Peisajovich, S.G., Constanzo, M., Andrews, B.J., Boone, C.M., Desveaux, D. and Guttman, D.S. (2019). Identifying Pseudomonas syringae Type III Secreted Effector Function via a Yeast Genomic Screen. *G3: Genes, Genomes, Genetics*, [online] 9(2), pp.535–547. Available at: https://www.g3journal.org/content/9/2/535.abstract [Accessed 20 May 2020].

Mansfield, J., Genin, S., Magori, S., Citovsky, V., Sriariyanum, M., Ronald, P., Dow, M., Verdier, V., Beer, S.V., Machado, M.A., Toth, I., Salmond, G. and Foster, G.D. (2012). Top 10 plant pathogenic bacteria in molecular plant pathology. *Molecular Plant Pathology*, [online] 13(6), pp.614–629. Available at: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1364-3703.2012.00804.x [Accessed 21 Apr. 2019].

Morel, A., Guinard, J., Lonjon, F., Sujeeun, L., Barberis, P., Genin, S., Vailleau, F., Daunay, M.-C., Dintinger, J., Poussier, S., Peeters, N. and Wicker, E. (2018). The eggplant AG91-25 recognizes the Type III-secreted effector RipAX2 to trigger resistance to bacterial wilt (Ralstonia solanacearum species complex). *Molecular Plant Pathology*, 19(11), pp.2459–2472.

Munkvold, K.R., Martin, M.E., Bronstein, P.A. and Collmer, A. (2008). A Survey of the Pseudomonas syringae pv. tomato DC3000 Type III Secretion System Effector Repertoire Reveals Several Effectors That Are Deleterious When Expressed in Saccharomyces cerevisiae. *Molecular Plant-Microbe Interactions®*, 21(4), pp.490–502.

Nahar, K., Matsumoto, I., Taguchi, F., Inagaki, Y., Yamamoto, M., Toyoda, K., Shiraishi, T., Ichinose, Y. and Mukaihara, T. (2013). Ralstonia solanacearumtype III secretion system effector Rip36 induces a hypersensitive response in the nonhost wild eggplantSolanum torvum. *Molecular Plant Pathology*, 15(3), pp.297–303.

Parsons, A.B., Brost, R.L., Ding, H., Li, Z., Zhang, C., Sheikh, B., Brown, G.W., Kane, P.M., Hughes, T.R. and Boone, C. (2003). Integration of chemical-genetic and genetic interaction data links bioactive compounds to cellular target pathways. *Nature Biotechnology*, 22(1), pp.62–69.

Peeters, N., Guidot, A., Vailleau, F. and Valls, M. (2013). Ralstonia solanacearum, a widespread bacterial plant pathogen in the post-genomic era. *Molecular Plant Pathology*, 14(7), pp.651–662.

Popa, C., Li, L., Gil, S., Tatjer, L., Hashii, K., Tabuchi, M., Coll, N.S., Ariño, J. and Valls, M. (2016). The effector AWR5 from the plant pathogen Ralstonia solanacearum is an inhibitor of the TOR signalling pathway. *Scientific Reports*, 6(1).

Proft, M., Pascual-Ahuir, A., de Nadal, E., Ariño, J., Serrano, R. and Posas, F. (2001). Regulation of the Sko1 transcriptional repressor by the Hog1 MAP kinase in response to osmotic stress. *The EMBO Journal*, 20(5), pp.1123–1133.

Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S., Arlat, M., Billault, A., Brottier, P., Camus, J.C., Cattolico, L., Chandler, M., Choisne, N., Claudel-Renard, C., Cunnac, S., Demange, N., Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schiex, T., Siguier, P., Thébault, P., Whalen, M., Wincker, P., Levy, M., Weissenbach, J. and

Boucher, C.A. (2002). Genome sequence of the plant pathogen Ralstonia solanacearum. *Nature*, 415(6871), pp.497–502.

Salomon, D., Bosis, E., Dar, D., Nachman, I. and Sessa, G. (2012a). Expression of Pseudomonas syringae type III effectors in yeast under stress conditions reveals that HopX1 attenuates activation of the high osmolarity glycerol MAP kinase pathway. *Microbiology*, 158(11), pp.2859–2869.

Salomon, D., Bosis, E., Dar, D., Nachman, I. and Sessa, G. (2012b). Expression of Pseudomonas syringae type III effectors in yeast under stress conditions reveals that HopX1 attenuates activation of the high osmolarity glycerol MAP kinase pathway. *Microbiology*, 158(11), pp.2859–2869.

Salomon, D., Dar, D., Sreeramulu, S. and Sessa, G. (2011). Expression of Xanthomonas campestris pv. vesicatoria Type III Effectors in Yeast Affects Cell Growth and Viability. *Molecular Plant-Microbe Interactions®*, 24(3), pp.305–314.

Siggers, K.A. and Lesser, C.F. (2008). The Yeast Saccharomyces cerevisiae: A Versatile Model System for the Identification and Characterization of Bacterial Virulence Proteins. *Cell Host & Microbe*, 4(1), pp.8–15.

Slagowski, N.L., Kramer, R.W., Morrison, M.F., LaBaer, J. and Lesser, C.F. (2008). A Functional Genomic Yeast Screen to Identify Pathogenic Bacterial Proteins. *PLoS Pathogens*, 4(1), p.e9.

Sreelatha, A., Bennett, T.L., Carpinone, E.M., O'Brien, K.M., Jordan, K.D., Burdette, D.L., Orth, K. and Starai, V.J. (2014). Vibrio effector protein VopQ inhibits fusion of V-ATPase–containing membranes. *Proceedings of the National Academy of Sciences*, 112(1), pp.100–105.

Sreelatha, A., Bennett, T.L., Zheng, H., Jiang, Q.-X., Orth, K. and Starai, V.J. (2013). Vibrio effector protein, VopQ, forms a lysosomal gated channel that disrupts host ion homeostasis and autophagic flux. *Proceedings of the National Academy of Sciences*, 110(28), pp.11559–11564.

Wang, J., Yeckel, G., Kandoth, P.K., Wasala, L., Hussey, R.S., Davis, E.L., Baum, T.J. and Mitchum, M.G. (2020). Targeted suppression of soybean BAG6-induced cell death in yeast by soybean cyst nematode effectors. *Molecular Plant Pathology*, 21(9), pp.1227–1239.

A4.5 Figures

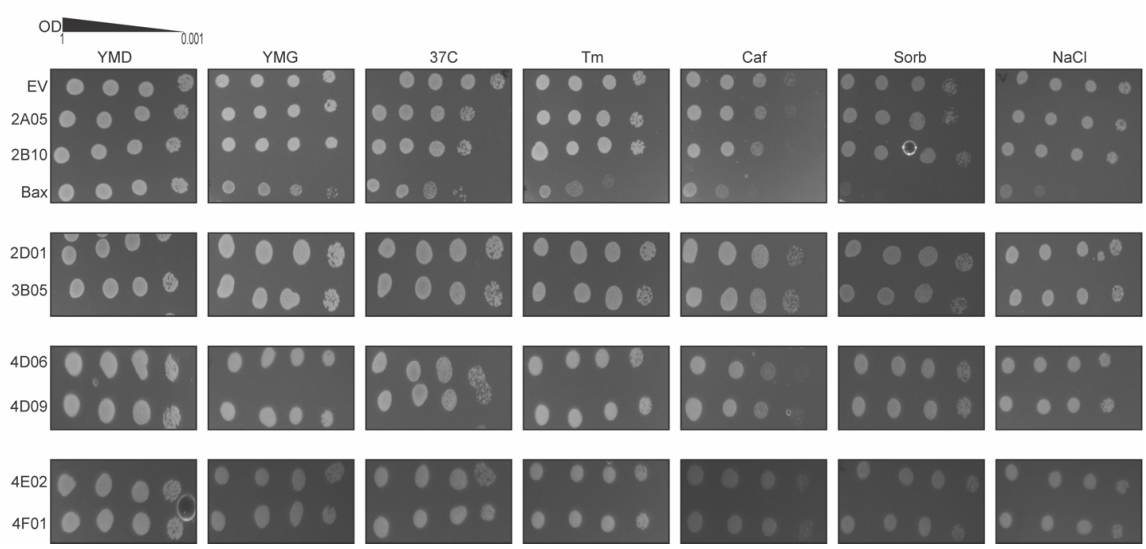


Figure 1: Expression of HG type 0 nematode effectors in pSH221 in the presence or absence of additional stressors: YMD (yeast minimal media with 2% glucose), YMG (yeast minimal media with 2% galactose and 1% raffinose), caffeine (Caf:7 mM in YMG), tunicamycin (Tm: 0.12 μg/mL in YMG), sorbitol (sorb: 1 M in YMG), and NaCl (0.5 M in YMG). Images taken after five days growth at 30°C or 37°C.

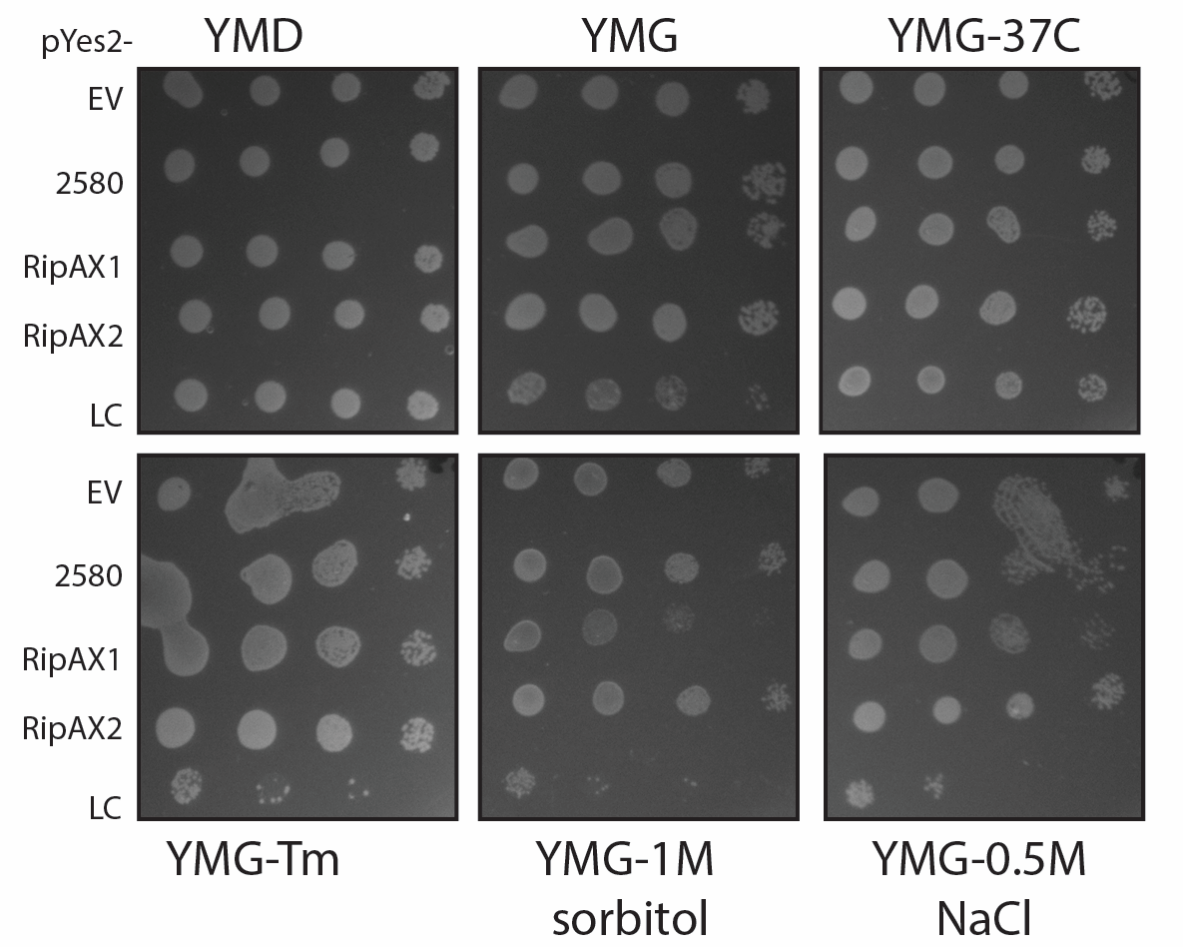


Figure 2: Similar to Figure 1 except expression of *R. solanacearum* RipAX1 and RipAX2 effectors, and no caffeine treatment. LC=α-SNAP$_{Rhg1}$LC

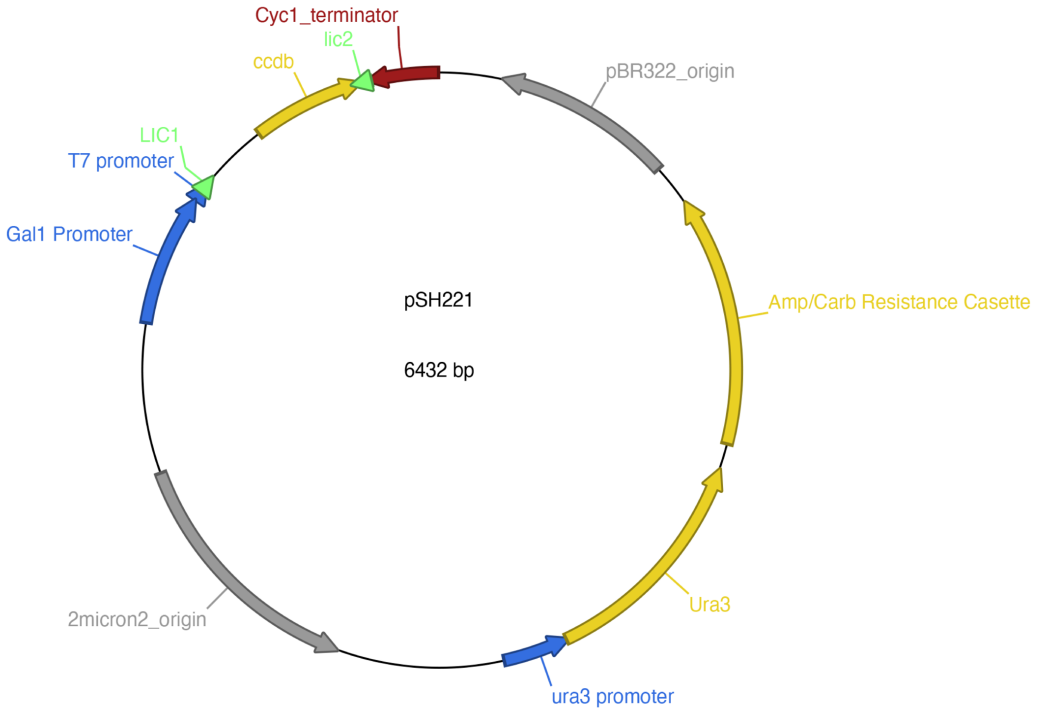
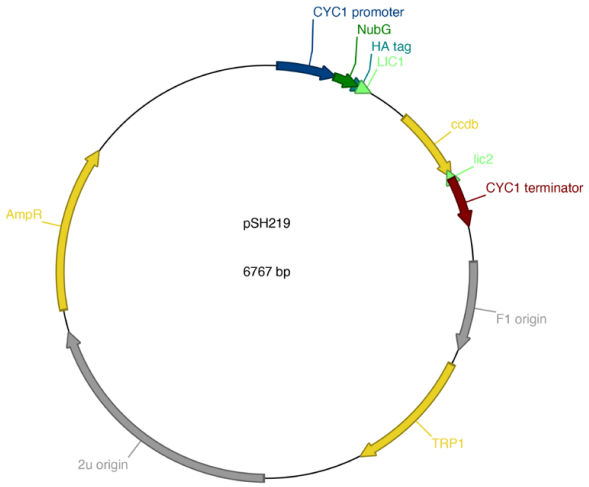
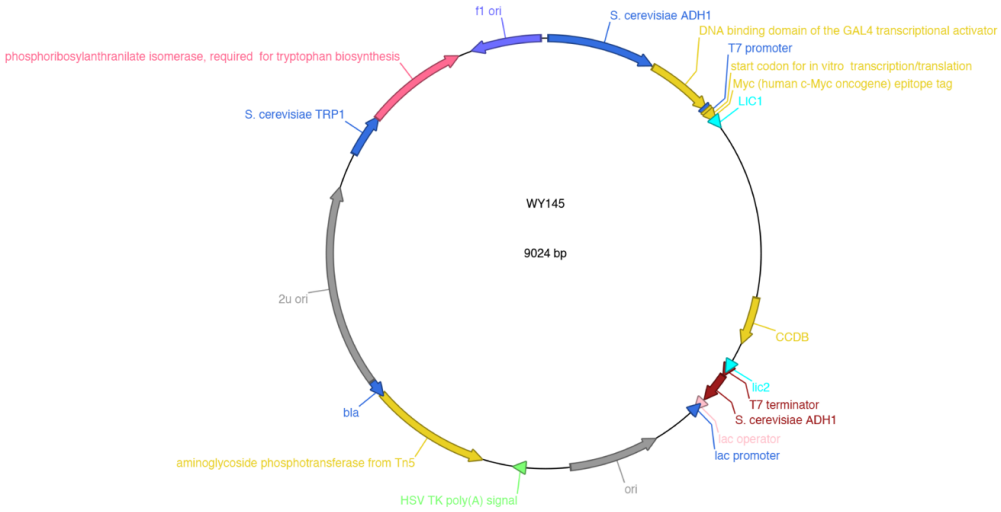


Figure 3: pSH221 plasmid used for inducible expression of effectors, BAX, or α-SNAP in yeast. Cloning is done by ligase independent cloning. Bacterial selection is ampicillin, and yeast selection is uracil auxotrophy.

A


CYC1 promoter
NubG
HA tag
LIC1
ccdb
lic2
CYC1 terminator
pSH219
6767 bp
AmpR
F1 origin
TRP1
2u origin

B


f1 ori
S. cerevisiae ADH1
DNA binding domain of the GAL4 transcriptional activator
T7 promoter
start codon for in vitro transcription/translation
Myc (human c-Myc oncogene) epitope tag
LIC1
phosphoribosylanthranilate isomerase, required for tryptophan biosynthesis
S. cerevisiae TRP1
WY145
9024 bp
2u ori
CCDB
lic2
T7 terminator
S. cerevisiae ADH1
lac operator
lac promoter
bla
aminoglycoside phosphotransferase from Tn5
HSV TK poly(A) signal
ori

C

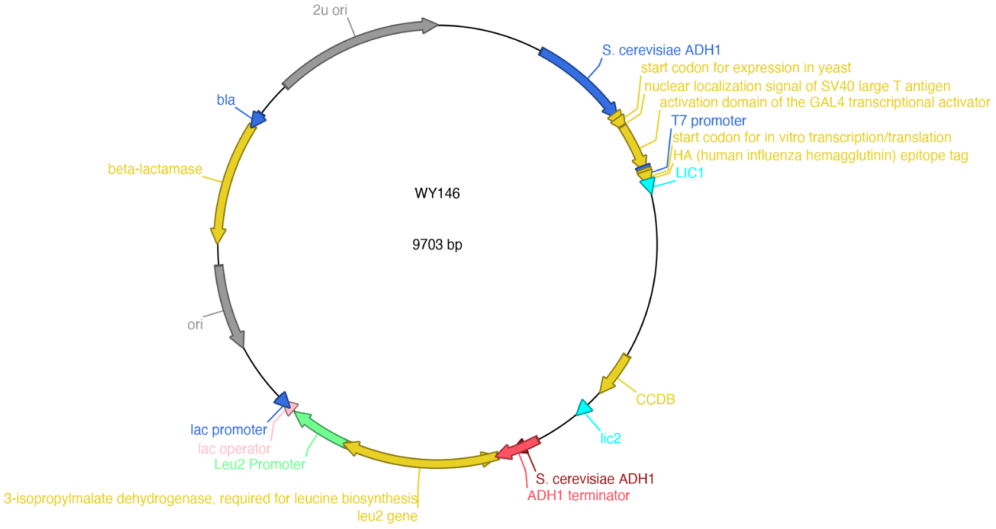

2u ori
S. cerevisiae ADH1
start codon for expression in yeast
nuclear localization signal of SV40 large T antigen
activation domain of the GAL4 transcriptional activator
T7 promoter
start codon for in vitro transcription/translation
HA (human influenza hemagglutinin) epitope tag
LIC1
bla
beta-lactamase
WY146
9703 bp
ori
CCDB
lac promoter
lac operator
Leu2 Promoter
lic2
S. cerevisiae ADH1
ADH1 terminator
3-isopropylmalate dehydrogenase, required for leucine biosynthesis
leu2 gene

Figure 4: Yeast two-hybrid vectors. (A) pSH219 is used for cloning effectors to be fused with an N-terminal portion of ubiquitin for split-ubiquitin experiments by ligase independent cloning. Bacterial selection is ampicillin, and yeast selection is tryptophan auxotrophy. (B) WY145 is used for N-terminally fusing effectors or *Rhg1*-encoded genes to the Gal4-DNA binding domain by ligase independent cloning, for yeast two-hybrid experiments. Bacterial selection is kanamycin, and yeast selection is tryptophan auxotrophy. (C) WY146 is used for N-terminally fusing effectors or *Rhg1*-encoded genes to the Gal4-activation domain by ligase independent cloning. Bacterial selection is ampicillin, and yeast selection is leucine auxotrphy.

Appendix 5: Investigating *Rhg1* in Other Legume Species.

Derrick Grunwald, Jessica Peterson, Andrew Bent

For common bean, I performed all extractions, cloning, and sequencing. For perennial *Glycine* species, I was assisted in extractions, PCR, cloning and sequencing by Jessica Peterson.

A5.1 Introduction

Common bean (*Phaseolus vulgaris*) is an important grain crop belonging to the same family as *Glycine max*, *Fabaceae*, and thought to have been domesticated from two gene pools: Andean and Mesoamerican (Singh et al., 1991; McClean et al., 2004). Indeed, soybean and common bean are both produced in the same regions and the top-producing soybean producing states: North Dakota, Minnesota, and Michigan constitute 60% of the total common bean production in the United States (usdrybeans.com). Furthermore, many of the same pathogens that are destructive to soybean production also can affect common bean, most notably soybean cyst nematode (Poromanto et al., 2010). Due to the overlap in growing regions and timing of planting (spring or early summer), common bean may provide an inoculum source of SCN for soybeans and vice versa. Therefore, control of SCN in common bean provides a mechanism to similarly control field levels of SCN in areas where both are planted.

The genetics of disease resistance in common bean has been empowered by recent advances in next generation sequencing technologies. Genome wide association studies have been applied to common bean resistance to both bacterial blight and SCN (Shi et al., 2011; Wen et al., 2019). Indeed, in a recent study by Wen et al. (2019), QTLs associated with SCN HG type 2.5.7 and HG type 1.2.3.5.6.7 resistance were found on chromosomes 1, 7 and 9. Interestingly, the chromosome 1 QTL is proximal to a region that has synteny with the *Rhg1* locus in soybean. As soybean and common bean are not interfertile, presence of *Rhg1* within common bean suggests it is a shared-derived locus, and that *Rhg1*-mediated resistance might be a derived trait.

Perennial *Glycine* species constitute a set of 27 species indigenous to Australia and other pacific islands, not thought to be interfertile with *Glycine max*. Due to their unique adaptation to diverse agronomic and climatic conditions, there has been much research on the genetics of disease resistance and abiotic stress tolerance within the perennial *Glycine* species. In fact, resistance to diverse pathogens including alfalfa mosaic virus, bean pod mottle virus, brown spot, powdery mildew, soybean sudden death syndrome, sclerotinia stem rot, and soybean cyst nematode have all been found within accessions of the perennial *Glycine* species (Mignucci and Chamberlain, 1978; Burdon and Marshall, 1981; Lim and Hymowitz, 1987; Hartman et al., 1992; Horlock et al., 1997; Riggs et al., 1998; Hartman et al., 2000; Zheng et al., 2005; Bauer et al., 2007; Wen et al., 2017; Herman et al., 2019).

For SCN in particular, there have been a handful of large-scale evaluations of perennial *Glycine* accessions for resistance. In 1998, Riggs et al. reported resistance of 9 perennial species to HG types 2, 1.2, and 1.3. In a later comprehensive screen, Bauer et al., tested nearly 500

perennial accessions and found that over half had resistance to HG type 0 nematode populations (2007). Later studies also found that many perennial accessions were additionally resistant to HG type 1.2.3.4.5.6.7, and 2.5.7 nematode populations. Like common bean, whether this resistance is mediated by *Rhg1* and, if so, whether this resistance is a shared-derived trait, are open questions.

A5.2 Results and Discussion

A polymorphic α-SNAP is not associated with *Rhg1* in select *P. vulgaris* accessions

Emboldened by the discovery of *Rhg1* within the annual wild-relative of soybean, we wanted to investigate the presence of *Rhg1* within other legume species: common bean *P. vulgaris* demonstrates a range of susceptibility to soybean cyst nematode (Wen et al., 2019). Importantly, *P. vulgaris* is not thought to be interfertile with domesticated soybean, suggesting that presence of *Rhg1*, and particularly multi-copy *Rhg1* would be a derived trait in the *Fabaceae* family. Indeed, in a Genome wide association study performed by Wen et al. (2019), a syntenic *Rhg1* block on chromosome 1 was found to correlate with resistance to HG type 2.5.7 and HG type 1.2.3.5.6.7 in *P. vulgaris* across the Mesoamerican and Andean gene pools. As a polymorphic α-SNAP is known to correlate with SCN resistance, we first sought to characterize the α-SNAP gene on chromosome 1 (PHAVU_001G247900). Sequencing of select common bean accessions from the Mesoamerican and Iranian gene pools that are moderately susceptible (PI 313373), resistant (PI 533428 and PI 201296), or untested (PI 416648) to HG type 0 nematode populations revealed that there was no difference between the accessions at the C-terminal region of the α-SNAP, the region associated with SCN resistance (Figure 1). This suggests that,

at least in these varieties, a polymorphic α-SNAP is not responsible for driving HG type 0 resistance.

Interestingly, in the Wen et al. (2019) study, the same GWAS experiment did not reveal significant association between the ortholog of $SHMT_{Rhg4}$ and SCN resistance. As above, this suggests that resistance, in particular to HG type 2.5.7 nematode populations, in common bean, does not rely on sources known to confer resistance in domesticated soybean. Future work may be directed to elucidating the mechanism of resistance in common bean, in particular as it pertains to other loci thought to associate with resistance on chromosomes 7 and 9 (Wen et al., 2019). These loci may have homologous regions in soybean that, through genetic manipulation or through natural variation, may similarly be utilized to confer resistance in these populations. Such discovery would expand the library of available resistance QTLs for breeders to incorporate into their breeding programs.

It should be noted that we looked only at a handful of common bean accessions from one gene pool. It is possible that there are accessions in either the Andean (which was not checked) or remaining Mesoamerican gene pool which either have multi-copy *Rhg1* and/or a polymorphic α-SNAP. Developing primer sets that amplify the *Rhg1* in common bean will go a long way to determining the potential origins of multi-copy *Rhg1* within the *Fabaceae* family. As it stands, there remain ~300 common bean accessions that are resistant to either HG type 2.5.7 or HG type 1.2.3.5.6.7 nematode populations which are available to screen in such a way (Wen et al., 2019).

Resistance to HG type 2.5.7 nematodes in select perennial *Glycine* species is not associated with multi-copy *Rhg1* We next sought to determine whether resistance to HG type 2.5.7 nematode populations in perennial *Glycine* species is associated with multi-copy *Rhg1*. In particular, we screened the 16 accessions described in Herman et al. (2019) as being HG type 2.5.7 resistant and which constituted 4 different species (*G. tomentella*, *G. clandestina*, *G. argyria*, and *G. dolichocarpa*) for presence of multi-copy *Rhg1*. Using the primer sets described in chapter 4, we found that none of the resistant accessions gave amplicons indicating presence of tandem duplication of *Rhg1*, whereas there were amplicons indicating presence of at least one copy of *Rhg1* (Figure 2). This suggests that in the perennial *Glycine* species we screened, single copy *Rhg1* is present, and its duplication is likely not associated with HG type 2.5.7 resistance.

We further sought to determine whether an aberrant α-SNAP, associated with multi-copy *Rhg1*, was present in the putatively single-copy perennial *Glycine* species. We tested 4 accessions for presence of an aberrant α-SNAP using our custom antibodies, which recognize α-$SNAP_{Rhg1}$WT as well as α-$SNAP_{Rhg1}$HC, α-$SNAP_{Rhg1}$LC, and α-$SNAP_{Rhg1}$GSM. Testing each of the accessions revealed that none had an α-SNAP recognized by antibodies raised against α-$SNAP_{Rhg1}$HC or α-$SNAP_{Rhg1}$LC, but each did have an α-SNAP recognized by α-$SNAP_{Rhg1}$WT (Figure 3).

The results presented here are reminiscent of what we and others have observed in *Glycine soja*. In particular, PI 468916 lacks multi-copy *Rhg1* but has strong resistance to HG type 2.5.7 nematode populations through an unknown mechanism that relies on the QTLs *cqSCN-006* and *cqSCN-007* (Kabelka et al., 2005; Kabelka et al., 2006; Kim and Diers, 2013; Yu and

Diers, 2017). Furthermore, in our own studies in *G. soja*, we found that multi-copy *Rhg1* was associated with resistance to HG types 0 and 2.5.7 nematode populations but was likely not the only contributing locus (chapter 4). Altogether, the results presented here are consistent with wild populations of crop species, here soybean, have alternative means of conferring resistance to pests and pathogens that remain to be exploited by conventional breeding or genetic modification approaches (Hyten et al., 2006; Liu et al., 2020).

Two questions that arise are: (1) is there a perennial *Glycine* species accession that does have multi-copy *Rhg1*? And (2) what is the mechanism of resistance in those perennial *Glycine* species that do not have multi-copy *Rhg1*? Comparisons between the genomes of wild perennial *Glycine* species and their domesticated relatives have revealed extensive synteny and collinearity between their chromosomes (Cannon et al., 2014; Varshney et al., 2013; Chang et al., 2014). Since these initial studies, a complete draft genome for *G. latifolia* has been developed (Liu et al., 2018). Such a draft genome may serve as a template to design more specific primer sets for *Rhg1* or other loci of interest, future RNAseq experiments, and validation of genes of interest through CRISPR/CAS9 experiments.

Resistance to HG type 2.5.7 nematodes in select perennial *Glycine* species is not associated with NSF_{RAN07} Although we were not able to find multi-copy *Rhg1* within any of the screened perennial *Glycine* species, we sought to determine if any of the accessions had NSF_{RAN07}. The lack of perennial *Glycine* species within the SoySNP50K dataset necessitates the use of PCR to screen for NSF_{RAN07} rather than using the previously described SNP marker (Song et al., 2013; Song et al., 205; Bayless et al., 2018). Using the primer sets described in chapters 2 and 4, we

screened through the 16 perennial *Glycine* accessions (Figure 3). Doing so, we found that the majority of the accessions gave an amplicon for NSF_{Ch07}. Interestingly, PI 446996 and PI 509468 both gave amplicons suggesting that these accessions might contain NSF_{RAN07}. However, these results were not able to be verified when preparing a separate seed stock.

We decided to validate that PI 509468 did not have the R_4Q polymorphism by cloning and sequencing the first exon of *Glyma.07G195900* from this accession and comparing to NSF_{Ch07} and NSF_{RAN07} and the NSF_{Ch07} from a different perennial glycine accession, PI 509451. Sequencing of the fist exon revealed that PI 509468 did not have the NSF_{RAN07}-associated polymorphism. However, there was a polymorphism ($H_{48}R$) in PI 509468 (Figure 5). The functional consequences, if any, of this polymorphism, as well as cloning the entirety of the gene to determine if there are additional polymorphisms can be evaluated.

Combined with the observations from common bean above, the results presented in this appendix suggest that, barring additional screening, multi-copy *Rhg1* and potentially NSF_{RAN07} arose uniquely in the annual *Glycine* species: *G. soja* and *G. max*. This would be consistent with the hypothesis that soybean was domesticated from a Chinese *G. soja* population, as perennial *Glycine* species are both not interfertile with *G. soja* and are endemic to Australia and Papua New Guinea (Zhou et al., 2015b; Kofsky et al., 2018). Future work might be directed to determine whether there are additional transitional forms of *Rhg1* or NSF_{RAN07} within the annual *G. soja*. While we speculate to have found a transitional *Rhg1* in chapter 4, additional NSF forms remain to be discovered. Discovery of such NSFs or the lack of additional forms would both provide powerful insights into the evolution of *Rhg1* and resistance to SCN.

A5.3 Methods

Plant materials Seeds from 4 common bean accessions (*Phaseolus vulgaris*) and perennial *Glycine* species were obtained from the USDA germplasm collection (GRIN). For perennial *Glycine* species, there were 11 *Glycine tomentella* (PI 339657, PI 441008, PI 446996, PI 446998, 483219, PI 483227, PI 505235, PI 505286, PI 509501, PI 563892, PI 537064), 2 from *G. clandestina* (PI 509457, 509468), 2 from *G. argyria* (PI 5050151, PI 509451) and one from *G. dolichocarpa* (PI 441005). Common bean seeds were germinated on wet filter paper until a radicle appeared. The seed was then transferred to soil and allowed to grow in a 16-hour photoperiod until sufficient tissue could be acquired for extraction.

DNA Extraction DNA was extracted from common bean leaf tissue essentially as described in Cook et al., 2012. DNA was extracted from perennial *Glycine* seeds using the CTAB method, described in Cook et al., 2012, with important modifications. Seed material was first ground in a mortar and pestle into a fine powder. The powder was then transferred to a 2.0 mL tube and 1 mL of extraction buffer [2% CTAB (w/v), 100 mM Tris•HCl (pH = 8.0), 20 mM EDTA (pH = 8.0), 1.4 M NaCl, 1% (w/v) PVP, 3 mM β-mercaptoethanol] was added. The mixture was then incubated at 65°C for 45 minutes. After 45 minutes, an equal volume of chloroform : isoamyl alcohol (96:4) was added, and the mixture was then centrifuged at 5000 g for 10 minutes at 4°C. The supernatant was then transferred to a fresh 2.0 mL tube, and 0.5 volumes of ammonium acetate (7.5 M) and 2.5 volumes of 100% ethanol were added. The mixture was allowed to incubate overnight at -20°C before being centrifuged briefly to collect the pellet. The pellet was then resuspended in TE, to which 0.1 volumes of sodium acetate (3M, pH = 5.2) and

2.5 volumes of 100% ethanol were added. The mix was then incubated at -20°C for 1 hour before being spun down at 5000 g for 10 minutes at 4°C. The pellet was then resuspended in TE and quantified using a spectrophotometer.

Immunoblotting and antibodies. Polyclonal rabbit antibodies raised against *Rhg1* α-SNAP$_{LC}$, *Rhg1* α-SNAP$_{HC}$ and α-SNAP$_{WT}$ were previously generated and validated in Bayless *et al.*, 2016. Tissue preparation and immunoblots were performed essentially as in Bayless *et al.*, 2016; Bayless *et al.*, 2018. Essentially, leaf or root tissue was flash frozen in liquid N_2, massed, them homogenized using a PowerLyzer24 (Qiagen) for three to four cycles of 10 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% (vol/vol) Triton X-100, 10% (vol/vol) glycerol, 1/100 Sigma protease inhibitor cocktail] was then added in a 3:1 volume to mass ratio. Lysates were then centrifuged at 7000 g and 7500 g for 10 minutes. Bradford assays were then performed on each supernatant and equal amount of OD_{595} total protein were loaded in each sample lane of an SDS/PAGE gel. Immunoblots for α-SNAP and NSF 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 1:1000. Secondary horseradish peroxidase conjugated goat anti-rabbit IgG (Sigma) was added at 1:10,000 and incubated for 1 hour at room temperature on a platform shaker followed by five washes with TBS-T. Chemiluminescence signal detection was performed with SuperSignal West Dura chemiluminescent substrate (Thermo Scientific) and developed using a Chemi Doc MP chemiluminescent imager (Bio-Rad).

Vector construction *Glyma.07G195900* was amplified from genomic DNA using Kapa HiFi polymerase (Roche) and placed directly under the control of the soybean ubiquitin promoter and nopaline synthase terminator in pBlueScript using Gibson assembly (Gibson et al., 2009), and subsequently sequenced verified.

Primers used in this appendix

5’ Rhg1 Fw: TGACTAAGCAGCAAGCACAGA

5’ Rhg1 Rev: TCTTTCCGTGTACGGTCGTC

3’ Rhg1 Fw: AGAATACGTGGAGGCACAGC

3’ Rhg1Rev: CCATGCGTTACGATGCGATG

Bridge Fw-1: TTTAGCCTGCTCCTCACAAATTC

Bridge Rev-1: TTGGAGAATATGCTCTCGGTTGT

NSF_{Ch07} Detect Fw: CAACACGCCCGCGAGCGAC

NSF_{RAN07} Detect Fw: CTACACGCCCGCGAACGAC

NSF Detect Rev: CTCACTTGTACGGAATCACCGG

PHAVU_001G247900 Fw: GTAAAATAGGAATACTTTACGTGG

PHAVU_001G247900 Rev: CCTTGTTTCCGAATGAACATATCGA

NSF_{Ch07} 5’ Genomic Gibson Fw: TTGTTGACTCGACAGGCTCGTAAGTCGTGTTTATAGCC

NSF_{Ch07} 5’ Genomic Gibson Rev: GGTCGAATTCGCCCTTTTCAAGCCAGTAGAACAGCATCAAT

A5.4 References

Bauer, S., Hymowitz, T. and Noel, G. (2007). Soybean cyst nematode resistance derived from Glycine tomentella in amphiploid (G. max X G. tomentella) hybrid lines. *Nematropica*, 37(2), pp.277–284.

Bayless, A.M., Zapotocny, R.W., Grunwald, D.J., Amundson, K.K., Diers, B.W. and Bent, A.F. (2018). An atypical N-ethylmaleimide sensitive factor enables the viability of nematode-resistant Rhg1 soybeans. *Proceedings of the National Academy of Sciences*, 115(19), pp.E4512–E4521.

Burdon, J. and Marshall, D. (1981). Evaluation of Australian native species of Glycine for resistance to soybean Rust. *Plant Disease*, 65(1), pp.44–45.

Cannon, S.B., May, G.D. and Jackson, S.A. (2009). Three Sequenced Legume Genomes and Many Crop Species: Rich Opportunities for Translational Genomics. *Plant Physiology*, 151(3), pp.970–977.

Chang, S., Thurber, C.S., Brown, P.J., Hartman, G.L., Lambert, K.N. and Domier, L.L. (2014). Comparative Mapping of the Wild Perennial Glycine latifolia and Soybean (G. max) Reveals Extensive Chromosome Rearrangements in the Genus Glycine. *PLoS ONE*, 9(6), p.e99427.

Cook, D.E., Lee, T.G., Guo, X., Melito, S., Wang, K., Bayless, A.M., Wang, J., Hughes, T.J., Willis, D.K., Clemente, T.E., Diers, B.W., Jiang, J., Hudson, M.E. and Bent, A.F. (2012). Copy Number Variation of Multiple Genes at Rhg1 Mediates Nematode Resistance in Soybean. *Science*, 338(6111), pp.1206–1209.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A. and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, [online] 6(5), pp.343–345. Available at: https://www.nature.com/articles/nmeth.1318.pdf.

Hartman, G., Wang, T. and Hymowitz, T. (1992). Sources of resistance to soybean rust in perennial Glycine species. *Plant Disease*, 76(4), pp.396–399.

Hartman, G.L., Gardner, M.E., Hymowitz, T. and Naidoo, G.C. (2000). Evaluation of Perennial Glycine Species for Resistance to Soybean Fungal Pathogens That Cause Sclerotinia Stem Rot and Sudden Death Syndrome. *Crop Science*, 40(2), pp.545–549.

Herman, T.K., Han, J., Singh, R.J., Domier, L.L. and Hartman, G.L. (2020). Evaluation of wild perennial Glycine species for resistance to soybean cyst nematode and soybean rust. *Plant Breeding*, 139(5).

Horlock, C., Teakle, D. and Jones, R. (1997). Natural infection of the native pasture legumes, Glycine latifola, by alfalfa mosaic virus in Queensland. *Australian Plant Pathology*, 26, pp.115–116.

Hyten, D.L., Song, Q., Zhu, Y., Choi, I.-Y., Nelson, R.L., Costa, J.M., Specht, J.E., Shoemaker, R.C. and Cregan, P.B. (2006). Impacts of genetic bottlenecks on soybean genome diversity. *Proceedings of the National Academy of Sciences*, 103(45), pp.16666–16671.

Kabelka, E.A., Carlson, S.R. and Diers, B.W. (2005). Localization of Two Loci that Confer Resistance to Soybean Cyst Nematode from Glycine soja PI 468916. *Crop Science*, 45(6), pp.2473–2481.

Kabelka, E.A., Carlson, S.R. and Diers, B.W. (2006). Glycine soja PI 468916 SCN Resistance Loci's Associated Effects on Soybean Seed Yield and Other Agronomic Traits. *Crop Science*, 46(2), pp.622–629.

Kim, M. and Diers, B.W. (2013). Fine Mapping of the SCN Resistance QTL and from PI 468916. *Crop Science*, 53(3), p.775.

Kofsky, J., Zhang, H. and Song, B.-H. (2018). The Untapped Genetic Reservoir: The Past, Current, and Future Applications of the Wild Soybean (Glycine soja). *Frontiers in Plant Science*, 9.

Lim, S.M. and Hymowitz, T. (1987). Reactions of Perennial Wild Species of Genus Glycine to Septoria glycines. *Plant Disease*, 71(10), p.891.

Liu, Q., Chang, S., Hartman, G.L. and Domier, L.L. (2018). Assembly and annotation of a draft genome sequence for Glycine latifolia, a perennial wild relative of soybean. *The Plant Journal*, 95(1), pp.71–85.

Liu, Y., Du, H., Li, P., Shen, Y., Peng, H., Liu, S., Zhou, G.-A., Zhang, H., Liu, Z., Shi, M., Huang, X., Li, Y., Zhang, M., Wang, Z., Zhu, B., Han, B., Liang, C. and Tian, Z. (2020). Pan-Genome of Wild and Cultivated Soybeans. *Cell*, 182(1), pp.162-176.e13.

Mignucci, J. and Chamberlain, D. (1978). Interactions of Microsphaera diffusa with Soybeans and Other Legumes. *Phytopathology*, 68(2), p.169.

Poromarto, S.H., Nelson, B.D. and Goswami, R.S. (2010). Effect of Soybean Cyst Nematode on Growth of Dry Bean in the Field. *Plant Disease*, 94(11), pp.1299–1304.

Riggs, R., Wang, S., Singh, R. and Hymowitz, T. (1998). Possible transfer of resistance to Heterodera glycines from Glycine tomentella to Glycine max. *Supplement to the Journal of Nematology*, 30(4S), pp.547–552.

Shi, C., Navabi, A. and Yu, K. (2011). Association mapping of common bacterial blight resistance QTL in Ontario bean breeding populations. *BMC Plant Biology*, 11(1), p.52.

Singh, S., Gepts, P. and Debouck, D. (1991). Races of Common Bean (Phaseolus vulgaris, Fabaceae). *Economic Botany*, 45, pp.379–396.

Song, Q., Hyten, D.L., Jia, G., Quigley, C.V., Fickus, E.W., Nelson, R.L. and Cregan, P.B. (2013). Development and Evaluation of SoySNP50K, a High-Density Genotyping Array for Soybean. *PLoS ONE*, 8(1), p.e54985.

Song, Q., Hyten, D.L., Jia, G., Quigley, C.V., Fickus, E.W., Nelson, R.L. and Cregan, P.B. (2015). Fingerprinting Soybean Germplasm and Its Utility in Genomic Research. *G3: Genes|Genomes|Genetics*, 5(10), pp.1999–2006.

Thomas Stalker, H., Charles Brummer, E. and Wison, R. (2004). Legume Crop Genomics.

Usadrybeans.com. (2020). [online] Available at: http://usadrybeans.com [Accessed 23 Nov. 2020].

Varshney, R.K., Song, C., Saxena, R.K., Azam, S., Yu, S., Sharpe, A.G., Cannon, S., Baek, J., Rosen, B.D., Tar'an, B., Millan, T., Zhang, X., Ramsay, L.D., Iwata, A., Wang, Y., Nelson, W., Farmer, A.D., Gaur, P.M., Soderlund, C., Penmetsa, R.V., Xu, C., Bharti, A.K., He, W., Winter, P., Zhao, S., Hane, J.K., Carrasquilla-Garcia, N., Condie, J.A., Upadhyaya, H.D., Luo, M.-C., Thudi, M., Gowda, C.L.L., Singh, N.P., Lichtenzveig, J., Gali, K.K., Rubio, J., Nadarajan, N., Dolezel, J., Bansal, K.C., Xu, X., Edwards, D., Zhang, G., Kahl, G., Gil, J., Singh, K.B., Datta, S.K., Jackson, S.A., Wang, J. and Cook, D.R. (2013). Draft genome sequence of chickpea (Cicer arietinum) provides a resource for trait improvement. *Nature Biotechnology*, 31(3), pp.240–246.

Wen, L., Chang, H.-X., Brown, P.J., Domier, L.L. and Hartman, G.L. (2019). Genome-wide association and genomic prediction identifies soybean cyst nematode resistance in common bean including a syntenic region to soybean Rhg1 locus. *Horticulture Research*, 6(1).

Wen, L., Yuan, C., Herman, T.K. and Hartman, G.L. (2017). Accessions of Perennial Glycine Species With Resistance to Multiple Types of Soybean Cyst Nematode (Heterodera glycines). *Plant Disease*, 101(7), pp.1201–1206.

Yu, N. and Diers, B.W. (2017). Fine mapping of the SCN resistance QTL cqSCN-006 and cqSCN-007 from Glycine soja PI 468916. *Euphytica*, 213(2).

Zheng, C., Chen, P., Hymowitz, T., Wickizer, S. and Gergerich, R. (2005). Evaluation of Glycine species for resistance to Bean pod mottle virus. *Crop Protection*, 24(1), pp.49–56.

Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, W., Yu, Y., Shu, L., Zhao, Y., Ma, Y., Fang, C., Shen, Y., Liu, T., Li, C., Li, Q., Wu, M., Wang, M., Wu, Y., Dong, Y., Wan, W., Wang, X., Ding, Z., Gao, Y., Xiang, H., Zhu, B., Lee, S.-H., Wang, W. and Tian, Z. (2015). Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nature Biotechnology*, 33(4), pp.408–414.

A5.5 Figures

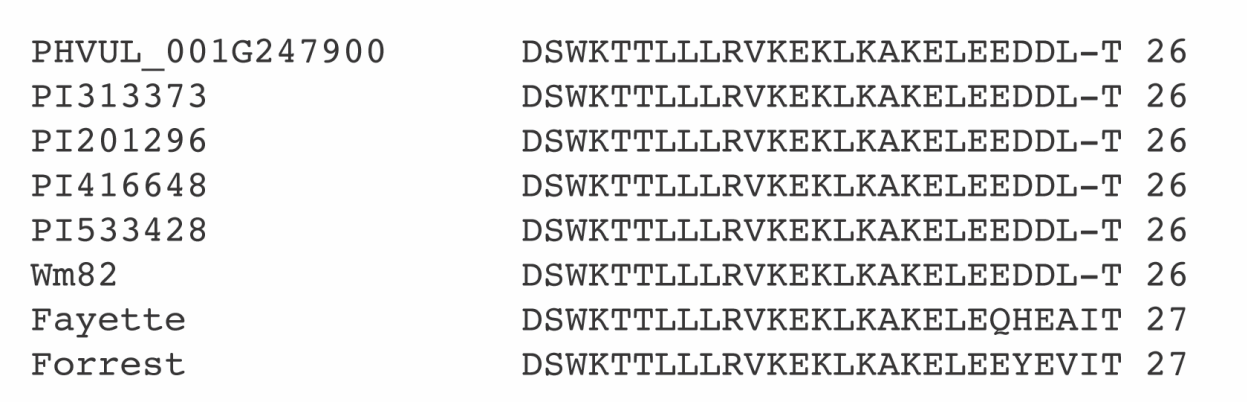


Figure 1: Alignment of C-terminal region of α-SNAP (PHAVU_001G247900) between resistant (PI 533428 and PI 201296), moderately susceptible (PI 313373), or untested (PI 416648) common bean accessions as well as resistant and susceptible *Glycine max*.

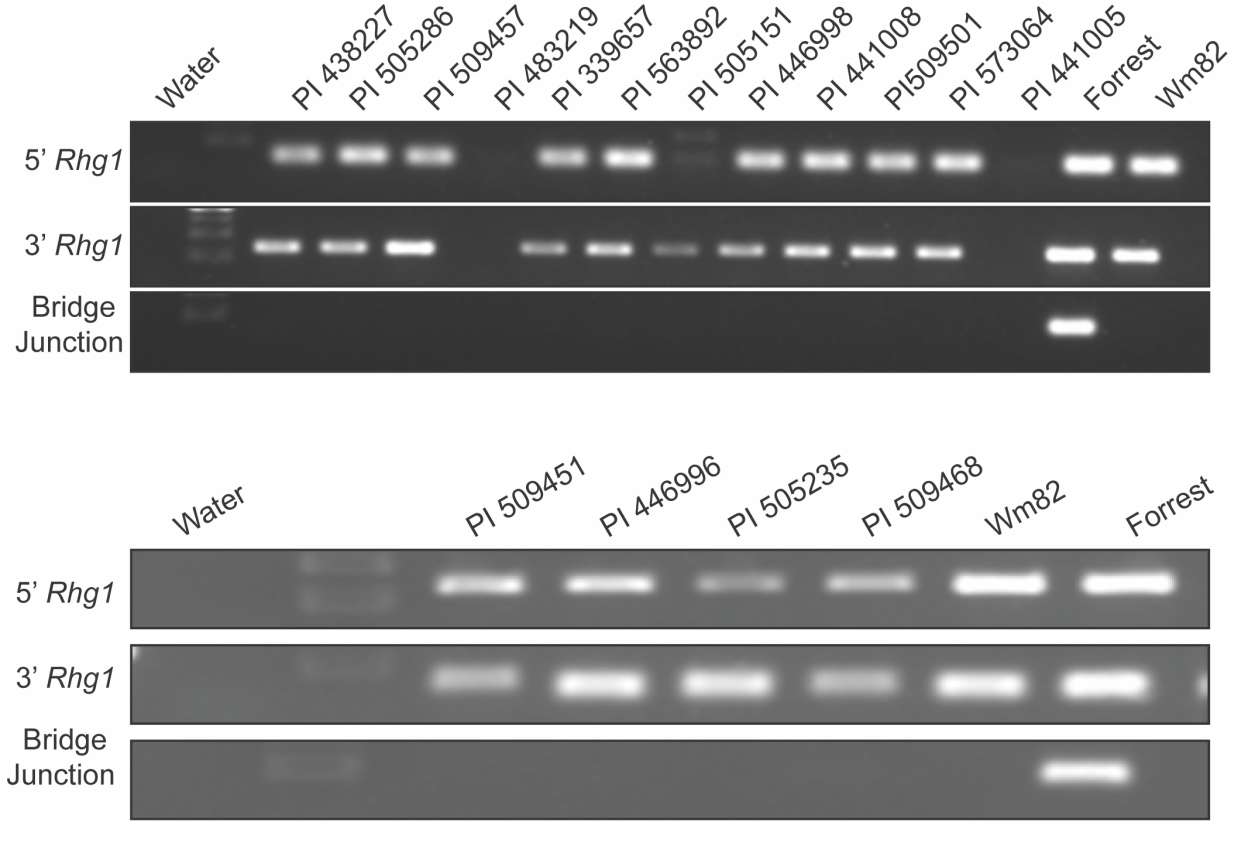


Figure 2: Results of PCR using primers specific for the 5' and 3' portions of the *Rhg1* locus common to both single and multi-copy lines. An outward-facing primer set was used to detect tandem duplication of the *Rhg1* locus.

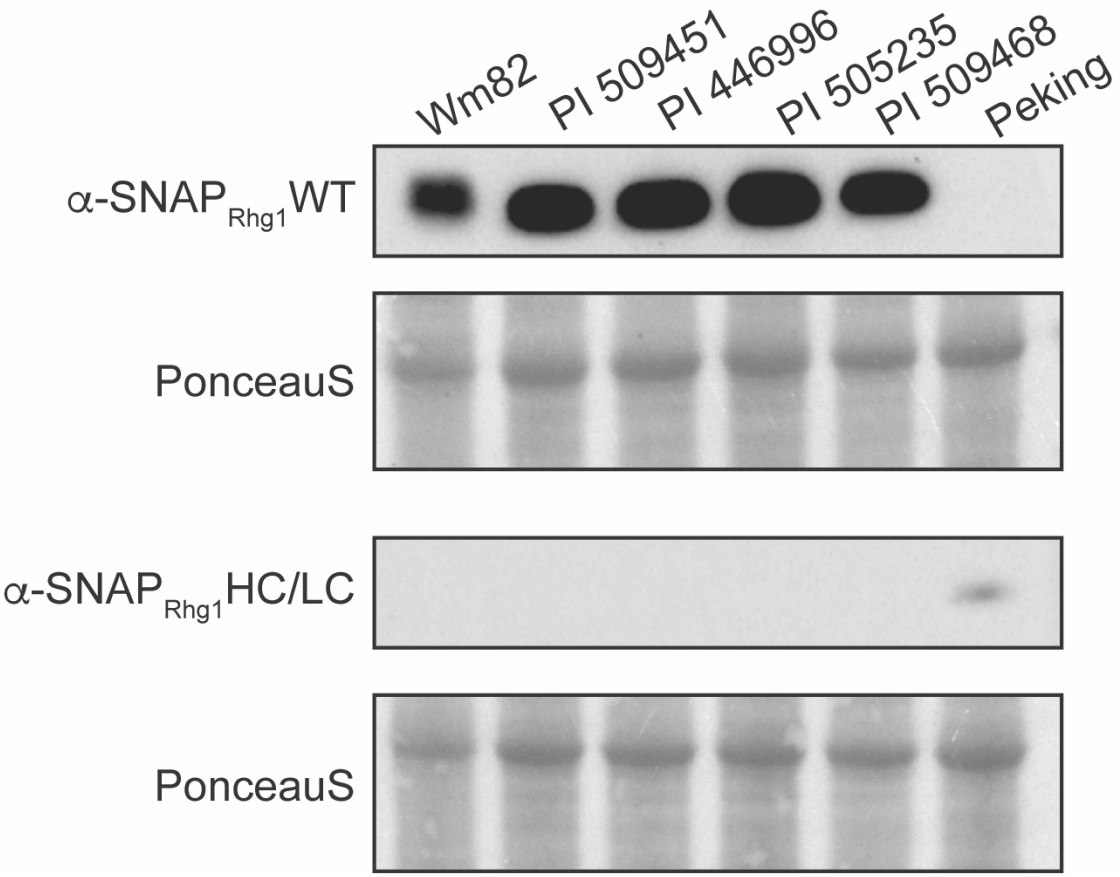


Figure 3: Immunoblots showing presence of α-SNAP in select perennial *Glycine* accessions and *Glycine max* accessions Williams 82 (containing α-$SNAP_{Rhg1}$WT) and Peking (containing α-$SNAP_{Rhg1}$LC) using antibodies raised against α-$SNAP_{Rhg1}$WT, α-$SNAP_{Rhg1}$HC, or α-$SNAP_{Rhg1}$LC. PonceauS shown as loading control.

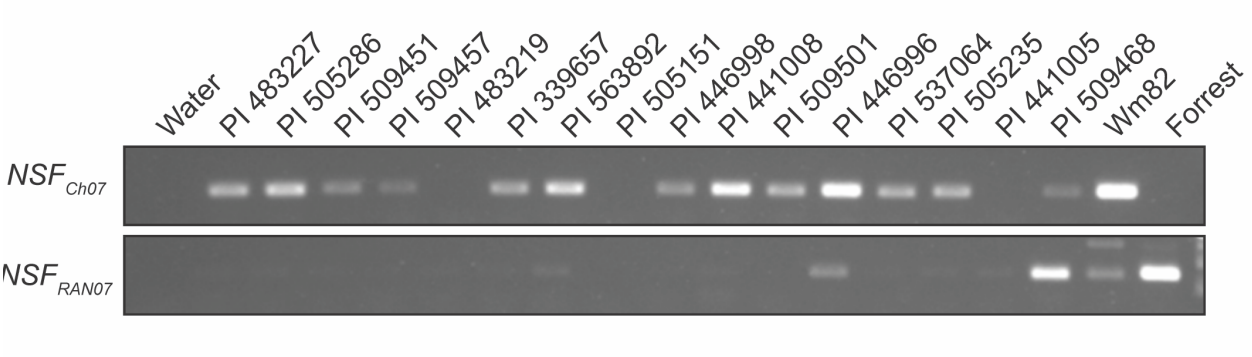


Figure 4: Screening for the presence of the R_4Q polymorphism (top) in perennial *Glycine* species relative to R_4 (NSF_{Ch07}; bottom).

```
RAN07          MASQFGLSSSSSSASSMRVTYTPANDLALTNLAFCSPSDLRNFAVPGHNNLYLAAVADSF
Wm_82          MASRFGLSSSSSSASSMRVTNTPASDLALTNLAFCSPSDLRNFAVPGHNNLYLAAVADSF
PI_509451      MASRFGLSSSSSSASSMRVTNTPASDLALTNLAFCSPSDLRNFAVPGHNNLYLAAVADSF
PI_505468      MASRFGLSSSSSSASSMRVTNTPASDLALTNLAFCSPSDLRNFAVPGRNNLYLAAVADSF
```

Figure 5: Alignment of the first exon of NSF (*Glyma.07G195900*) from NSF_{Ch07}, RAN07 and from the perennial *Glycine* species PI 509451 (which did not give an R_4Q amplicon) and PI 505468 (which gave an R_4Q amplicon). Polymorphisms unique to RAN07 marked in Red; polymorphisms unique to PI 505468 ($H_{48}R$) denoted in purple.