

APPLICATION OF CRISPR/CAS9-MEDIATED GENOME EDITING
FOR STUDYING SOYBEAN RESISTANCE TO SOYBEAN CYST NEMATODE

A Thesis
presented to
the Faculty of the Graduate School
at the University of Missouri-Columbia

In Partial Fulfillment
of the Requirements for the Degree
Master of Science

by
JINGWEN KANG
Dr. Melissa G. Mitchum, Thesis Advisor

December 2016

The undersigned, appointed by the dean of the Graduate School,
have examined the Thesis entitled

APPLICATION OF CRISPR/CAS9-MEDIATED GENOME EDITING FOR
STUDYING SOYBEAN RESISTANCE TO SOYBEAN CYST NEMATODE

Presented by Jingwen Kang a candidate for the degree of Master of Science and hereby
certify that, in their opinion, it is worthy of acceptance

Dr. Melissa G. Mitchum

Dr. Emmanuel Liscum

Dr. Anne L. McKendry

To my beloved parents.

Acknowledgments

I would first like to thank my thesis advisor Dr. Melissa Mitchum. The door to her office was always open whenever I ran into a trouble spot or had a question about my research or writing. She consistently allowed this paper to be my own work, but steered me in the right the direction whenever she thought I needed it. I would also like to thank the experts in our lab, especially Dr. Pramod Kandoth, who started mentoring me since my senior year back in the year of 2013. I would never be able to accomplish this research project without his help. Also, my sincere appreciation to all the other lab members. Without their consistent encouragement and advice, my project could not have been successfully conducted.

Table of Contents

Acknowledgements	ii
List of Illustrations	iv
Project Summary	v
Chapter 1: Literature Review	1
Overview of soybean cyst nematode	1
SCN management.....	2
Figure 1. 1. Cartoon diagram of the life cycle of SCN.....	2
SCN resistance in soybean	3
Genetic understanding of soybean resistance to SCN	4
Role of SHMT in SCN resistance.....	5
Overview of CRISPR/Cas9	7
Application of CRISPR/Cas9 in plants	9
Rationale.....	11
Chapter 2: Application of CRISPR/Cas9-Mediated Genome Editing for Studying Soybean Resistance to Soybean Cyst Nematode	13
Abstract	13
Introduction	14
Materials and Methods	17
Plant and nematode material	17
Bacterial strains	17
Cloning of <i>GmSHMT05</i>	17
Guide RNA design.....	18
Single gRNA plasmid construction	18
Dual gRNA plasmid construction.....	20
<i>Agrobacterium rhizogenes</i> transformation	22
Generation of soybean hairy roots.....	22
Genomic DNA extraction from hairy roots	22
Sequencing screening for modification in <i>GmSHMT08</i>	23
Off-targeting analysis	23
PCR screening for deletions in <i>GmSHMT08</i>	24
Nematode infection assay.....	24
Results	26
Application of CRISPR/Cas9 to knock out <i>GmSHMT08</i> in soybean hairy roots	26
Evaluation of CRISPR/Cas9 <i>GmSHMT08</i> soybean hairy roots in nematode infection	28
Using CRISPR/Cas9 to test <i>GmSHMT08</i> for a role in soybean basal resistance	29
Analysis of <i>GmSHMT05</i> in soybean	30
Discussion	34
Acknowledgements	57
Chapter 3: Summary and Future Perspectives	57
References	67

List of Illustrations

Tables	41
Table 2. 1. Summary of two types of CRISPR/Cas9 constructs targeting <i>GmSHMT08</i>	41
Table 2. 2. Nucleotide deletion analysis.....	42
Table 2. 3. Nucleotide and amino acid comparison between Forrest type and Essex type <i>GmSHMT08</i>	43
Table 2. 4. Summary of single gRNA construct modification efficiency on <i>GmSHMT05</i>	44
Figures	45
Figure 2. 1. Vector used for single gRNA construct.	45
Figure 2. 2. Positions of gRNAs on Forrest type <i>GmSHMT08</i>	46
Figure 2. 3. p201GCas9 targeting of <i>GmSHMT08</i> in soybean hairy roots.....	47
Figure 2. 4. Vector used to induce <i>GmSHMT08</i> deletions in hairy roots..	48
Figure 2. 5. Gel images of nucleotide deletions.	49
Figure 2. 6. Nucleotide deletion illustration and sequencing results.....	50
Figure 2. 7. Infection assay of EXF67 SHMT8-1 and SHMT8-4.....	52
Figure 2. 8. Infection assay of EXF67 SHMT8-14..	53
Figure 2. 9. Infection assay of EXF63 SHMT8-14..	54
Figure 2. 10. Amino acid alignment of <i>GmSHMT05</i> , Essex type <i>GmSHMT08</i> , and Forrest type <i>GmSHMT08</i>	55
Figure 2. 11. Infection assay of EXF67 and EXF63 SHMT5.	56
Appendix	63
Figure A.1. cDNA alignment of Forrest type <i>GmSHMT08</i> with <i>GmSHMT05</i>	63
Figure A.2. cDNA alignment of Essex type <i>GmSHMT08</i> with <i>GmSHMT05</i>	65

Project Summary

Soybean is one of the world's most important crops, providing billions of people with food, energy resources, and industrial materials. The value of soybean has increased to 41 billion dollars in the United States. Soybean cyst nematode (*Heterodera glycines*; SCN) has become one of the most economically important pathogens of soybean in the world. SCN causes more than 1.2 billion dollars in yield losses annually to soybean farmers in the US alone. The main strategy for management of SCN is planting resistant cultivars. Most of the resistant cultivars on the market are derived from plant introductions (PI) 88788, 548402 (Peking) and 437654. Currently, the molecular basis of soybean resistance to SCN is not fully elucidated. Since the discovery of the first quantitative trait loci (QTL) for resistance to *Heterodera glycines* (*Rhg*) in the early 1960s, researchers have been working to understand the molecular basis of SCN resistance in soybean. *Rhg1* and *Rhg4* are two major QTL conferring resistance to SCN. The PI 88788 source of resistance requires *Rhg1*, whereas Peking and PI 437654 resistance is bigenic requiring both *Rhg1* and *Rhg4*. At the *Rhg4* locus, a gene encoding a serine hydroxymethyltransferase (SHMT) has been confirmed to play a role in resistance to SCN. SHMT is ubiquitous in nature. This enzyme functions in the simultaneous interconversion of serine to glycine and tetrahydrofolate to 5, 10-methylenetetrahydrofolate. However, the pathways and mechanisms leading to SCN resistance remain to be elucidated. Although a combination of reverse genetic methods including RNAi (RNA interference), TILLING (Targeting Induced Local Lesions in Genomes), and VIGS (Virus-Induced Gene Silencing) have been used effectively to study SCN resistance genes in soybean, limitations such as off-target effects, incomplete

silencing, and background mutations can complicate analysis. More recent genome editing technologies have become appealing for studies of gene function in soybean. A new, simpler genome editing method called bacterial type II CRISPR (clustered regularly interspaced short palindromic repeats /Cas9 (CRISPR-associated) immune system has recently emerged. The CRISPR/Cas9 system only requires a Cas9 nuclease and a single guide RNA (sgRNA) to perform the genome editing. This project was designed to apply CRISPR/Cas9 methodology to test soybean genes for a role in SCN resistance, and to further characterize the function of Rhg4 (SHMT) in SCN resistance.

Chapter 1: Literature review

Overview of soybean cyst nematode

The soybean cyst nematode (*Heterodera glycines*; SCN) is a sedentary endoparasite that infects soybean (*Glycine Max*) roots. It was first identified in northeast China in 1899. The nematode was later found in Japan in 1915, in Korea in 1936, and Manchuria in 1938 (Li et al., 2011; Chang & Qiu, 2011). SCN was first identified in the United States in 1954 and has since spread to all soybean producing states (Riggs et al., 1977; Tylka & Marett, 2014). Currently, SCN is the number one pathogen of soybean and causes more than \$1.2 billion in yield loss in the United States annually (Koenning and Wrather., 2010). The life cycle of SCN is 25-30 days under optimal conditions and consists of five life stages punctuated by four molts (Figure 1.1). The hatched second-stage juvenile (J2) is the infective stage. J2 penetrate the plant cell using a hollow mouth structure called a stylet. Stylet-secreted effector proteins such as cellulases and pectinases facilitate penetration and migration through root tissues. Once the SCN juvenile reaches a single cell near the vasculature, stylet-secreted effectors are delivered into a single cell where they interact with plant signals to modify plant gene expression. These changes in gene expression lead to partial cell wall dissolution and the fusion of hundreds of adjacent cells to form a highly metabolically active feeding cell known as a syncytium. The syncytium functions as a nutrition sink for the nematode to obtain essential nutrients from the living host plant (Endo, 1986). Once the J2 nematode initiates feeding, a loss of somatic musculature ensues and the nematode remains sedentary for the remainder of its life cycle. Therefore, the success of SCN relies on the successful establishment of the

syncytium. After the fourth molt, the adult males regain their motility and leave the roots to fertilize females. The female body, which protrudes from the root, retains the majority of the eggs. A small number of eggs are secreted in a gelatinous matrix into the soil. Upon death forms, the characteristic lemon-shaped cysts are visible on the root surface.

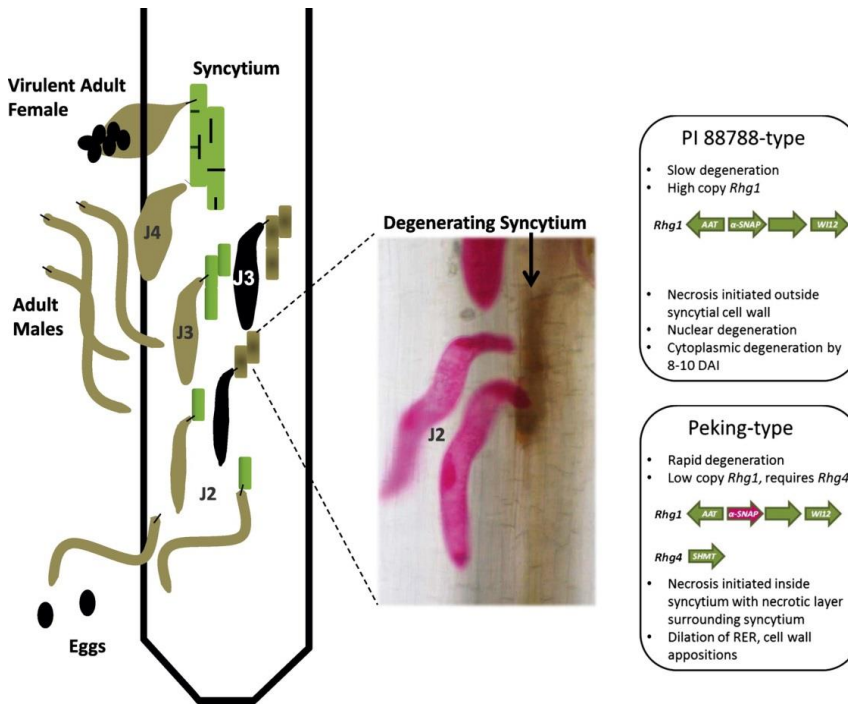


Figure 1. 1. Cartoon diagram of the life cycle of the soybean cyst nematode (SCN) in a Peking-type resistant soybean plant. The two black boxes highlight the features of two types of resistance to SCN Hg Type 0 (Kim et al. 1987, 2010b, 2012). Reproduced with permission from Mitchum, 2016.

SCN Management

Once SCN is introduced into a soybean field, it is impossible to eliminate it completely from the soil. Soybean eggs encased within cysts can remain viable for many years. Management of SCN relies heavily on planting SCN resistant soybeans and rotating with non-host crops to keep populations below an economic threshold.

SCN resistant soybean cultivars have been developed through conventional breeding using plant introductions (PIs) derived from the USDA germplasm collection as

sources of resistance. Although many PIs have been identified with SCN resistance, only a small number of these have been utilized for developing SCN resistant cultivars. Currently, 95% of SCN resistant soybeans are derived from a single source, PI 88788, and the remaining 5% from PI 548402 (Peking), PI 437654, or combinations of these. The overdependence on SCN resistant cultivars with the same genetic source of resistance has selected for nematodes that can overcome resistance thereby reducing the effectiveness (Mitchum, 2016). Recent statewide surveys have shown a significant shift towards virulence on SCN resistant cultivars containing resistance genes from PI 88788 (A. Howland and M. G. Mitchum, unpublished data). Therefore, novel SCN resistant cultivars are needed to manage different populations of SCN. Understanding the SCN resistance mechanism in soybean is essential to develop bioengineered soybean cultivars, as well as to maintain the high yield for soybean producers.

In addition to planting SCN resistant cultivars, it is recommended that soybean growers rotate planting of soybeans with non-host crops such as wheat or corn to avoid nematode population accumulation (Sasser & Grover, 1991). Planting non-host crops can stimulate some level of SCN hatching in the soil. The SCN juveniles penetrate roots of non-host crops, but cannot establish a feeding site, leading to a decrease of the SCN population for the next soybean growing season (Warnke et al., 2008).

SCN resistance in soybean

The resistance of soybean to SCN is based on a measure of nematode reproduction termed the female index (FI). The FI is calculated by taking an average cyst count across replicates of the test line and dividing by the average cyst count across replicates of the susceptible check and multiplying by 100. If the female index is less than

10% on a soybean line, the cultivar is regarded as resistant. In resistant soybean cultivars, infective juveniles are unable to establish a syncytium. Instead, the syncytium becomes necrotic and degenerates shortly after it is initiated in a manner similar to the hypersensitive response (HR) that leads to localized cell death in response to pathogens (Ghezzi et al., 1996). Since the syncytium provides SCN nutrition for growth and reproduction, the collapse of the syncytium will lead to the demise of nematode due to starvation. What differs among resistant soybean cultivars is the timing of the HR (Acedo et al., 1984). In Peking, syncytium degradation occurs around 2 days post-infection (Ghezzi et al., 1996). In contrast, in PI 209332, the HR response occurs at 8 to 10 days after SCN infection (Acedo et al., 1984). In both scenarios, the SCN penetrates the root but is unable to complete its life cycle. Whether the differences in the timing of HR have any biological impact on plants is still unknown.

Genetic understanding of soybean resistance to SCN

On a molecular level, how the SCN resistance mechanism is functioning in soybean is still not fully understood. The soybean genome is 1 Gb in size, which contains a total of 20 chromosomes that house all of the genes (Kim et al., 2010 Grant et al., 2010). Researchers have identified multiple quantitative trait loci (QTL) that are associated with SCN resistance in different PIs (Song et al., 2004). Two QTLs, *Rhg1* and *Rhg4* (“Rhg” for resistance to *Heterodera glycines*) are best characterized in SCN resistance. The *Rhg1* locus is on chromosome 18 and *Rhg4* is located on chromosome 8.

Rhg1 is required for resistance in all known sources of SCN resistance. In some soybean cultivars, such as Fayette (resistance derived from PI 88788) (Bernard et al., 1988), only *Rhg1* is required for resistance to SCN Hg type 0 (Race 3) (Concibido et al.,

2004). This type of resistance is referred to as “PI88788-type” resistance. In other cultivars such as Forrest (resistance derived from Peking) (Hartwig & Epps, 1973), resistance is bigenic requiring both *Rhg1* and *Rhg4* (Meksem et al., 2001). This type of resistance is referred to “Peking-type” resistance. Recently, SCN resistance (R) genes at the *Rhg1* (cv. Fayette) and *Rhg4* (cv. Forrest) loci were identified (Cook et al., 2012; Liu et al., 2012). *Rhg1* in Fayette was determined to be a 31kb segment containing a cluster of three dissimilar genes, each coding for an amino acid transporter, an α -SNAP (soluble NSF (N-ethylmaleimide-sensitive factor) attachment protein), and a WI12 (wound-inducible domain) protein duplicated in 10 tandem repeats (Cook et al., 2012). The high copy number leads to higher basal expression of the three genes, which contrasts with the susceptible cultivars that only have one copy of the gene cluster. A serine hydroxymethyltransferase (*SHMT*) gene was identified at *Rhg4* to play a role in Peking-type resistance (Liu et al., 2012). Although these genes have been identified on molecular level, the signaling pathways and underlying biochemical mechanisms to explain how they regulate SCN resistance remain to be elucidated.

Role of SHMT in SCN resistance

In Peking-type resistance, both *Rhg1* and *Rhg4* are required to confer resistance to SCN (Concibido et al. 2004; Meksem et al., 2001). Unlike PI 88788-type resistance, a low copy number of *Rhg1* is required when *Rhg4* is present which implies the expression of the *Rhg4* gene may have a more critical function in Peking-type SCN resistance (Acharya et al., 2015). The SHMT encoded by *Rhg4* in resistant cultivars differs from the susceptible form of SHMT by two amino acids that reside in the folate and pyridoxal-5-

phosphate (vitamin B6) binding sites (Liu et al., 2012). The above evidence indicates that SHMT may have gained a new enzymatic function during SCN parasitism.

SHMT is a pyridoxal phosphate-dependent enzyme that has been studied in numerous organisms. In mammals, two types of SHMT have been identified: mSHMT (found in mitochondria) and cSHMT (found in cytoplasm). In plants, SHMTs belong to multigene families with members localized in the cytoplasm, plastids, mitochondria, and nucleus (Zhang et al., 2010). A total of 117 *SHMT* genes have been identified in 18 representative plant genomes, including 14 copies of *SHMT* in soybean (Wu et al., 2016). The SHMT encoded by *Rhg4* in soybean is a predicted cytosolic enzyme (Liu et al., 2012). The major function of SHMT is to catalyze the reversible reaction of serine and tetrahydrofolate (THF) to glycine and 5-10 methylene tetrahydrofolate (MTHF). In consequence, the cellular one-carbon metabolism provides numerous downstream products that are critical for DNA synthesis, cellular methylation, and plant defense. Although the formation of the syncytium requires altered metabolic activities and increased DNA synthesis, whether the function of this SHMT is related to folate homeostasis has not been proven. Additionally, since soybean is an ancient polyploid (paleopolyploid), closely related copies of *SHMT* genes have appeared in the soybean genome through gene duplication (Schmutz et al., 2010). A recent study has shown the *Rhg4* gene on chromosome 8 (*GmSHMT08*) has a closely related homolog on chromosome 5 (*GmSHMT05*) (Wu et al., 2016). *GmSHMT08* and *GmSHMT05* have the same length and structure (Wu et al., 2016). These two genes were predicted to be the result of a whole genome duplication event that happened in 13 million years ago (Schmutz et al., 2010). Interestingly, through sequencing the two genes in 33 cultivated

and 68 wild soybeans, it was found that the SCN-resistant allele is absent from wild soybeans (Wu et al., 2016). These data suggest that the function of *GmSHMT08* in SCN resistance possibly emerged via artificial selection during the domestication of soybean (Wu et al., 2016). The function of *GmSHMT05* is still unknown.

Overview of CRISPR/Cas9

Recently, an alternative genome editing method called CRISPR (Clustered Regularly Interspaced Short Palindromic Repeat)/Cas (CRISPR-associated) system has become a major nuclease-based RNA-guided genome engineering research approach (Shen et al., 2014). The CRISPR/Cas system is widely found in bacteria and archaea and is an adaptive immune response to defend against invading viral and plasmid DNAs (Sorek et al., 2013; Terns & Terns, 2011). When a virus attacks, bacteria integrate the corresponding sequences of the invading DNA or RNA as short fragments/spacers into the cell's genome at the CRISPR locus. Since the DNA sequences on the CRISPR locus are separated by equal length spacer sequences, this 'DNA-repeat'-spacer-'DNA repeat' pattern gave way to the name of 'CRISPR'. When comparing different CRISPR loci, some common repeats were identified (Kunin et al, 2007). At the CRISPR loci, spacers interspace a cluster of *Cas* genes and a series of repeat sequences. Following transcription of the spacers to produce short fragments of CRISPR RNA (crRNA), CAS proteins use crRNA to match sequences with foreign invading genetic components of bacteria and destroy foreign DNA (Hsu et al., 2014).

The CRISPR/Cas9 used in this study is a type II CRISPR/Cas system that is best characterized in a wide range of bacterial and archaeal hosts compared to type I and type III systems (Hsu et al., 2013). The type II CRISPR/Cas9 system consists of a Cas9

nuclease and a single guide RNA (sgRNA) to target a DNA sequence (Belhaj et al, 2013). The Cas9 endonuclease, a large monomeric DNA nuclease comprising RuvC-like domain and HNH nuclease domain, is derived from the *Streptococcus progenies* type II CRISPR/Cas system (Xie & Yang, 2013). Originally, together with two short noncoding RNA molecules—crRNA and trans-activating crRNA (tracrRNA), Cas9 nuclease is guided to a specific DNA sequence and cleaves on both strands to induce a double strand break (DSB) that leads to DNA repair mechanisms through either non-homologous end-joining (NHEJ) or homologous end-joining (HDR). In most cases, plants repair DNA mainly by NHEJ, which results in unfaithful repairs to create small nucleotide deletions or insertions (Indels) (Podevin et al., 2013). Consequently, indels will introduce a frameshift mutation that impact gene function or result in a complete gene knock-out. To achieve successful Cas9 recognition on the target DNA sequence, a protospacer-adjacent motif (PAM) adjacent to the targeted DNA sequence is necessary. The -NGG is known as the PAM site. In order to target on different DNA sites, the Cas9 nuclease sequences from *Streptococcus progenies* remain the same in a CRISPR/Cas9 plasmid. Only the 20nt sgRNA needs to be re-designed. Based on the target gene sequence, two complementary oligonucleotides are designed to make a new sgRNA. In this way, the new sgRNA will allow the system targeting a different DNA sequence in the form of 5'-N(20)-NGG (Gasiunas et al., 2012).

The components of the CRISPR/Cas9 system are less complex than Zinc Finger Nuclease (ZFNs) or Transcription Activator-Like Effector Nuclease (TALENs) making construction of CRISPR/Cas9 constructs less labor intensive and relatively easier to design. Compared to ZFNs or TALENS, CRISPR/Cas9 does not require extra separate

diametric proteins for each specific target site (Bortesi & Fischer, 2015). Additionally, by introducing multiple gRNAs one can edit several genes simultaneously (Belhaj et al., 2013; Mao et al., 2013). This allows for precise knocking-out of redundant genes or closely-related pathways. Moreover, by targeting two sites on the gene of interest, large genomic deletions or inversions can be induced by introducing two DSBs (Li et al, 2013; Upadhyay et al., 2013; Zhou et al., 2014).

Application of CRISPR/Cas9 in plants

In recent years, CRISPR/Cas9 has been successfully applied in different model plants such as *Arabidopsis*, *Nicotiana benthamiana* (Li et al., 2013), and *Nicotiana tabacum* (Gao et al., 2015). Also, studies have applied CRISPR/Cas9 to genetically modify a diverse range of crops including wheat (Upadhyay et al., 2013), maize (Liang et al., 2014) and rice (Mao et al., 2013). A few of the studies are related with plant defense to pathogens. In wheat, for example, researchers have applied CRISPR/Cas9 to test if all three *MILDEW RESISTANCE LOCUS* alleles confer resistance to powdery mildew and successfully generated powdery mildew-resistant wheat plants (Wang et al., 2014).

While CRISPR/Cas9 has been used successfully for genome editing in plants, the biggest concern of CRISPR/Cas9 is often related to mutation efficiency. In *Arabidopsis* and rice, modification efficiencies up to 90% have been reported (Feng et al., 2014; Liang et al., 2014); however, the mutation frequency is often unpredictable due to different methods of transformation and the uniqueness of various plant species. For example, using PEG-based protoplast transformation in rice and in wheat, the mutagenesis efficiency was reported to be 15%-38% and 3%-8%, respectively (Shan et al., 2013; Xie & Yang, 2013). Agrobacterium infiltration in *N. benthamiana* leaves resulted in

mutagenesis efficiency of 2.7%-4.8%, whereas targeting endogenous genes using PEG-based protoplast transformation in *N. benthamiana* showed mutation efficiency up to 38% (Li et al., 2013). For soybean research, CRISPR/Cas9 has been applied to generate mutations in soybean hairy roots (Jacobs et al., 2015). Rates of mutation up to 95% were reported (Jacobs et al., 2015). Nevertheless, in another study applying CRISPR/Cas9 in soybean hairy roots, the modification rate was reported to be ~54% in 170 transgenic hairy roots (Cai et al., 2015). Additionally, targeting of the same gene using the same CRISPR/Cas9 backbones but different sgRNAs could result in variable mutation efficiencies (Cai et al., 2015).

Several factors influencing mutation efficiency should be considered. For example, the promoters driving sgRNA and components of sgRNA could play important roles in modification efficiency (Belhaj et al., 2015). For example, in soybean hairy roots, use of the Arabidopsis U6 promoter vs. the endogenous soybean U6 promoter to drive the gRNA cassette resulted in different mutation frequencies (Du et al., 2016). Whether or not changing gRNA promoters will improve the capacity of CRISPR/Cas9 in other plants still remains unknown (Cai et al., 2015). As for the sgRNA component, many sources provide the rankings of gRNAs according to various algorithms to evaluate the mutation efficiency, but why some gRNAs proved to have higher mutation rates than others is unknown. Still, researchers are trying to optimize the guidelines of predicting efficiency of gRNAs. In a recent study comparing different gRNAs targeting on the same gene in a human cell, the efficiency result showed that higher or lower GC contents were less effective in mutation compared to those gRNAs that have average GC content (Wang et al., 2014). Also, plants regenerate during growth and development, therefore, when the

mutation occurs often is unpredictable. Consequently, if mutations in different tissues occur independently, a chimeric plant consisting of cells that have different genotypes will form. Therefore, the mixture of wildtype, heterozygous, homozygous or biallelic loci could complicate the detection and accuracy of mutagenesis (Belhaj et al., 2015).

Nevertheless, researchers are investing great effort to investigate details of application of CRISPR/Cas9 in plants. Compared to traditional mutagenesis approaches, CRISPR/Cas9 still has relatively high mutation efficiency and low off-targets effects (Podevin et al., 2013). Thus, CRISPR/Cas9 shows promise to accelerate plant breeding through precise genome modification of homologous genes and multiple gene families simultaneously to improve polyploid crops.

Rationale

Currently, no transgenic or any genetically modified SCN resistant soybeans have been developed for commercial agricultural production. Nevertheless, biotechnology has been applied to identify and study SCN resistance genes in soybean. A combination of positional cloning and reverse genetic approaches including TILLING (Targeting Induced Local Lesions IN Genomes), VIGS (virus-induced gene silencing), and RNAi (RNA interference) confirmed that the *SHMT* gene was the gene for resistance at the *Rhg4* locus. However, these approaches have limitations. VIGs and RNAi are not only unable to knock out target genes on a specific chromosome, but also lack specificity of silencing. For both RNAi and VIGs, the minimum size gene fragment that can be used is approximately 300 bp to achieve efficient silencing. Therefore, while VIGS and RNAi demonstrated that *GmSHMT* played a role in SCN resistance, they could not distinguish between *GmSHMT08* and *GmSHMT05* (Liu et al., 2012). Using TILLING, two

GmSHMT08 mutants, Forrest 6266 (F6266) and F6756 were identified from an EMS-mutagenized population of soybean cv. Forrest. These two mutants exhibited increased susceptibility and ruled out a role for *GmSHMT05* in SCN resistance (Liu et al., 2012). However, TILLING is limited by background mutations that may interfere with the study of the target gene, and functional redundancy can make it difficult to elucidate gene functions.

This project focused on evaluating type II CRISPR/Cas9 methods to identify soybean genes with a role in SCN resistance by targeting *GmSHMT08* in a “proof-of-concept” study, as well as further characterize its function and that of *GmSHMT05* in basal resistance to SCN and root development.

Chapter 2: Application of CRISPR/Cas9-Mediated Genome Editing for Studying Soybean Resistance to Soybean Cyst Nematode

Abstract

Soybean cyst nematode (SCN) *Heterodera glycines* has become one of the most economically important pathogens of soybean in the world. The main strategy of SCN management is planting resistant soybean cultivars. The main sources of resistance include plant introductions (PI) PI 88788 and PI 548402 (Peking) from the USDA germplasm collection. *Rhg1* and *Rhg4* represent two major SCN resistance QTLs in these soybean PIs. Peking-type resistance requires both *Rhg1* and *Rhg4* for resistance, while PI 88788-type resistance only requires *Rhg1*. In Peking-type resistance, at the *Rhg4* locus on chromosome 8, a gene encoding an enzyme called serine hydroxymethyltransferase (*GmSHMT08*) has been confirmed to play a role in resistance to SCN. Additional copies of *SHMT* on different chromosomes have been identified, including a closely related *SHMT* gene on chromosome 5 (*GmSHMT05*), but the function of this gene in soybean remains to be determined. Although reverse genetic methods such as RNA interference (RNAi), Targeting Induced Local Lesions in Genomes (TILLING), and Virus-Induced Gene Silencing (VIGS) have been used effectively to study SCN resistance in soybean, limitations such as off-targeting, incomplete silencing, and background mutations can potentially complicate the analysis. In recent years, the novel genome editing clustered regularly interspaced short palindromic repeat (CRISPR)/ CRISPR-associated (Cas) system has shown great promise for precise genome editing to generate knockouts. In this study, type II CRISPR/Cas9 methods were evaluated as an approach to identify soybean

genes with a role in SCN resistance by targeting *GmSHMT08*, as well as further characterize its function and that of *GmSHMT05* in basal resistance and root development.

Introduction

Soybean cyst nematode (SCN) *Heterodera glycines* is a microscopic roundworm that causes more than \$1.2 billion in yield loss in the United States annually (Koenning & Wrather, 2010). Soybean producers manage SCN through planting resistant cultivars. The natural resistance currently found in more than 95% of SCN resistant soybean in the north central USA is derived from a single source, plant introduction (PI) 88788. Although other sources of resistance have been identified, few are widely available on the market. This lack of genetic diversity has led to the selection of SCN populations that can reproduce on resistant cultivars. A lack of information on SCN virulence genes and other factors playing roles in the plant-nematode interaction has complicated SCN management. Therefore, a better understanding of the resistance mechanism in soybean could help soybean growers better manage SCN and select proper SCN resistant cultivars for their fields.

Genetic mapping approaches have identified two major QTL for SCN resistance, *Rhg1* on chromosome 18 and *Rhg4* on chromosome 8. In some soybean cultivars, such as Fayette (resistance derived from PI 88788), only *Rhg1* is required for resistance to SCN HG type 0 (Race 3). This type of resistance is referred to as “PI 88788-type” resistance. In other cultivars such as Forrest (resistance derived from Peking), resistance is digenic requiring both *Rhg1* and *Rhg4* (Meksem et al., 2001). This type of resistance is referred to “Peking-type” resistance. Recently, SCN resistance (R) genes at the *Rhg1* (cv. Fayette)

and *Rhg4* (cv. Forrest) loci were identified (Cook et al., 2012; Liu et al., 2012). A cluster of three genes at the *Rhg1* locus tandemly repeated 10 times in the genome was shown to confer resistance to SCN in soybean cv. Fayette. These genes are Glyma18g02580 (encoding a predicted amino acid transporter), Glyma18g02590 (encoding a predicted α -soluble N-ethylmaleimide-sensitive factor attachment protein or SNAP vesicle-trafficking protein), and Glyma18g02610 (WI12, encoding a wound-inducible protein without a predicted function) (Cook et al., 2012). In the Peking-type of resistance, three tandemly repeated copies of the *Rhg1* gene cluster are present and it remains unclear whether all three genes are contributing to SCN resistance. Of the three genes, the α -SNAP protein is the only protein within the cluster with amino acid polymorphisms between resistant genotypes. We recently cloned the *Rhg4* gene for SCN resistance on chromosome 8 in soybean cv. Forrest (Liu et al., 2012). *Rhg4* is a dominant resistance gene encoding a serine hydroxymethyltransferase (SHMT). Two amino acids differ between the SHMT in SCN-susceptible cv. Essex and SCN-resistant cv. Forrest.

The SHMT encoded by *Rhg4* in soybean is a predicted cytosolic enzyme (Liu et al., 2012). The major function of SHMT is to catalyze the reversible reaction of serine and tetrahydrofolate (THF) to glycine and 5-10 methylene tetrahydrofolate (MTHF) providing 1-C units critical for DNA synthesis, cellular methylation, and plant defense. A recent study has shown the *Rhg4* gene on chromosome 8 (*GmSHMT08*) has a closely related homolog on chromosome 5 (*GmSHMT05*) (Wu et al., 2016). This is often the case in soybean, an ancient tetraploid that now functions as a diploid (Singh and Hymowitz, 1988). *GmSHMT08* and *GmSHMT05* were found to have the same length and structure, which could be the result of a whole genome duplication event that happened 13 million

years ago (Schmutz et al., 2010). Sequencing of these two genes in 33 cultivated and 68 wild soybeans determined that the SCN-resistant allele is absent from wild soybeans (Wu et al., 2016). Thus, the function of *GmSHMT08* in SCN resistance appears to have emerged via artificial selection during the domestication of soybean (Wu et al., 2016). However, the function of *GmSHMT05* is still unknown.

A combination of reverse genetic methods including TILLING, VIGS, and RNAi were used to confirm a role for *GmSHMT08* in resistance to SCN. Each approach has its unique advantages, but also its own limitations. For example, TILLING requires a mutagenized population in a genetic background with the trait of interest and is labor intensive to both generate and screen. Due to the highly duplicated genome of soybean, the ability to identify mutants of interest can be masked by functional redundancy in the genome. VIGS and RNAi only result in gene silencing and can be limited by cross-silencing of closely related genes. In recent years, the novel genome editing tool CRISPR/Cas9 has shown promise for precise genome modification in a diverse range of organisms (Hsu et al., 2014). Several recent studies have reported the successful use of CRISPR/Cas9 in soybean (Du et al., 2016; Michno et al., 2015; Cai et al., 2015; Sun et al., 2015; Jacobs et al., 2015). Studies to elucidate the underlying mechanism of SCN resistance in soybean would benefit from the application of this new technology. Therefore, in this study we targeted the *SHMT* gene (*GmSHMT08*) which has been demonstrated to play a role in SCN resistance to test the utility of two different CRISPR/Cas9 systems and its application to our well-established soybean hairy root infection assay pipeline.

Materials and Methods

Plant and nematode material

Forrest (Hartwig & Epps, 1973), which harbors SCN resistance derived from Peking, requires both *Rhg1* and *Rhg4*. Essex (Smith & Camper, 1973) is a SCN susceptible cultivar. EXF67 (*rhg1_Frhg1_FRhg4_FRhg4_F*) is resistant because it carries Forrest resistance alleles at both *Rhg1* and *Rhg4*. EXF63 (*rhg1_Frhg1_FRhg4_ERhg4_E*) carries the Forrest *Rhg1* allele and the Essex *Rhg4* allele and is susceptible to SCN. Both EXF63 and EXF67 are amenable to *Agrobacterium rhizogenes* hairy root transformation. Soybean cyst nematode *Heterodera glycines* inbred population PA3 (HG type 0) used in this study was mass-selected on soybean cv. Williams 82 according to standard procedures at the University of Missouri.

Bacterial strains

Escherichia coli strain DH5 α was used for propagation of CRISPR/Cas9 constructs with certain antibiotics. For soybean hairy root transformation *Agrobacterium rhizogenes* strain K599 was used for transformation. *E.coli* was cultured in LB media at 37°C. *A. rhizogenes* was cultured in LB or YEP media in 28°C.

Cloning of *GmSHMT05*

The full length *GmSHMT05* coding sequence was PCR amplified from cDNA generated from RNA extracted from soybean seedling leaf tissue of cultivars Forrest, Essex, EXF63 and EXF67. The Qiagen RNeasy Plant Mini kit was used to extract total RNA according to the manufacturer's instructions. cDNA was synthesized using PrimeScript 1st strand cDNA synthesis kit. Sequences were amplified with two flanking

primers (SHMT5_seq_F and SHMT5_seq_R) positioned immediately upstream of the ATG and downstream of the stop codon. Two other primers (SHMT5 seq_F2 and SHMT5 seq_R2) positioned in the middle of the coding region were used to check the sequence at the beginning and the end of the coding regions. Amplified fragments were sequenced at the MU DNA core facility.

Primers	Sequence 5'-3'
SHMT5_seq_R	CTTTGTGCCGAACCTTTGAGTGACAC
SHMT5_seq_F	TGGGCACTCACATTAGCATTCAC
SHMT5 seq_F2	GGCTCTTGCCGTGGCGCT
SHMT5 seq_R2	CGGAGATCTTCTTGCCGCCG

Guide RNA design

To design a gRNA targeting *GmSHMT08* (Glyma08g11490.2), full-length cDNA information of soybean cultivar Forrest was used. Based on multiple previous studies, the gRNA sequence is in the form of GN19 (N=A,T,C or G) adjacent to a PAM (Protospacer Adjacent Motif) site which is NGG (N =A,T,C, or G). The websites (<http://www.genome.arizona.edu/crispr/>, <http://cbi.hzau.edu.cn/cgi-bin/CRISPR>) were used to predict potential endogenous CRISPR sites on Glyma08g11490.2. From the several predicted gRNAs, gRNA targeting sites with low off-targeting frequency were selected. The specificity of candidate gRNA sequences were further confirmed using Phytozome (<https://phytozome.jgi.doe.gov/>) and SoyKB (<http://soykb.org/>).

Single gRNA plasmid construction

pUC gRNA vector that has the protospacer sequences and the *Medicago truncatula* U6.6 promoter (a gift from Wayne Parrott, University of Georgia, Athens) was

used for cloning the sgRNAs under the U6.6 pol III promoter. For each novel sgRNA, specific forward and reverse primers with sgRNA target sequences and 15 bp 5' sequence that is homologous to the flanking pUC gRNA shuttle vector sequences were designed. A forward primer which spans the 5' end of the MtU6.6 promoter and a reverse primer spanning the 3' end of the protospacer sequence were also designed. The forward primer contains *Spe-I* restriction site and reverse primer has *I-Ppol* site, respectively. Overlap PCR was performed to engineer the sgRNA constructs in pUC gRNA shuttle vector. Briefly using pUC gRNA shuttle vector as the template, two linear single-strand DNA pieces containing complementary sgRNA sequences were synthesized through two sets of PCR using sgRNA primers and flanking primers. The second PCR using the flanking primers on the products of the first PCRs as template resulted in a linear double-stranded piece of DNA containing sgRNA sequences under *Medicago truncatula* U6.6 promoter, protospacer sequences, and *Spe-I* and *I-Ppol* restriction sites on the two ends. This PCR product was digested with *Spe-I* and *I-Ppol* and ligated with pUC gRNA shuttle vector. The sgRNA cassette was then subcloned into binary vector p201GCas9 (a gift from Wayne Parrott, University of Georgia, Athens), which carries the Cas9 nuclease gene sequences, to create the plasmid p201GCas9-GmSHMT08.

Primers used to introduce sgRNA to pUC gRNA Shuttle vector	
Target 1: AGGGAAGTCGCTGATAAGTG	
Rhg4SHMT CPR-F1	GAGGGAAGTCGCTGATAAGTGGTTTTAGAGCTAGA AATAGCAAGTT
Rhg4SHMT CPR-R1C	CACTTATCAGCGACTTCCCTCAAGCCTACTGGTTTCG CTTGAAG
Target 2: CCAGCCCTACTCCGGCTCCC	
Rhg4SHMT CPR-F2N	CCCTACTCCGGCTCCCGTTTTAGAGCTAGAAATAGC

	AAGTT
Rhg4SHMT CPR-R2C	GGAGCCGGAGTAGGGCTGGCAAGCCTACTGGTTCG CTTGAAG
Target 3: AGGCCAAAACTCATAATCTG	
Rhg4SHMT CPR-F3	GAGGCCAAAACTCATAATCTGGTTTTAGAGCTAGAA ATAGCAAGTT
Rhg4SHMT CPR-R3C	CAGATTATGAGTTTTGGCCTCAAGCCTACTGGTTCG CTTGAAG
Target 4: GCTTGCGGGGCCACGTGCG	
Rhg4SHMT CRP-F4	GCGGGGCCACGTGCGTTTTAGAGCTAGAAATAGC AAGTT
Rhg4SHMT CRP-R4	GCACGTGGGCCCGCAAGCAAGCCTACTGGTTCGCT TGAAG
Target 5 :GTCAACGTCCAGCCCTACTC	
Rhg4SHMT CRP-F5	ACGTCCAGCCCTACTCGTTTTAGAGCTAGAAATAGC AAGTT
Rhg4SHMT CRP-R5	AGTAGGGCTGGACGTTGACAAGCCTACTGGTTCGCT TGAAG
Primers used to synthesize sgRNA through PCR	
Spe gRNA F	AAAAC TAGTATGCCTATCTTATATGATCAATGAGG
Ippol gRNA R	TGCTACTCTCTTAAGGTAGCAAAAAAAGCACCGACT CGGTG
Primers used to check final vector construct by PCR	
gRNA F	ATATACTCTCTTAAGGTAGCATGCCTATCTTATATGA TCAATGAGG
gRNA R	TGCTACTCTCTTAAGGTAGCAAAAAAAGCACCGACT CGGTG
Cas9F6	GTACTGGTTGTGGCCAAAGTGG
nosTrev	TGATAATCATCGCAAGACC

Dual gRNA plasmid construction

35S-Cas9-SK as the CRISPR/Cas9 construct backbone and Atu6-26-SK as sgRNA template were gifts from Jian-Kang Zhu's lab. 35S-Cas9-SK contains the coding sequence of hSpCas9 with *KpnI* and *XhoI* on the two ends. The Cas9 cassette was sub

cloned into pBluescript SK+ vector. After sgRNA sequences were selected, complementary oligomers were designed to contain sgRNA sequences with *BbsI* restriction site at 5' ends. To anneal oligos, forward and reverse sequences were mixed and incubated for 10 minutes at 95°C followed by slow cooling to room temperature for 20 minutes. Each annealed double-stranded sgRNA sequence with *BbsI* sites were cloned into Atu6-26-SK vector digested with *BbsI*. The sgRNA expression cassettes in between *BamHI* and *SpeI*, *BamHI* and *PstI*, and the Cas9 expression cassette from 35S-Cas9-SK were subcloned into pcamGFP-CvMV-GWOX and the final vector was named as pCam-GWOX-SHMT8.

Primers to generate gRNA fragments	
Target 1: GCTTGCGGGGCCACGTGCG	
SHMT8 atusgRNA-4F	5' GATTGCTTGCGGGGCCACGTGCG 3'
SHMT8 atusgRNA-4R	5' AAACCGCACGTGGGCCCCGCAAGC 5'
Target 2: AGGGAAGTCGCTGATAAGTG	
SHMT8 atusgRNA-1F	5' GATTGAGGGAAGTCGCTGATAAGTG 3'
SHMT8 atusgRNA-1R	5' AAACCACTTATCAGCGACTTCCCTC 3'
Primers used for sequencing the Atu6-gRNA constructs in AtU6-26-SK	
M13R	5' CACAGGAAACAGCTATGAC 3'
SS42(sequencing)	5' TCCCAGGATTAGAATGATTAGG 3'
Primers used to check final vector construct by PCR	
SS42	5' TCCCAGGATTAGAATGATTAGG 3'
Cas9 F	5' GCCCAAGAGGAACAGAGTAAGC 3'
Cas9 R	5' CAGTTCGCCGGCAGAGGCCAGC 3'

***Agrobacterium rhizogenes* transformation**

The different sgRNA binary vector constructs were transformed into *Agrobacterium rhizogenes* strain K599 by freeze-thaw method (Weigel & Glazebrook, 2002).

Generation of soybean hairy roots

Soybean recombinant inbred line EXF67 was used for CRISPR silencing of the *GmSHMT08* gene. As controls, EXF67 and EXF63 (SCN susceptible) transformed with vector were used. Soybean seeds were surface sterilized in chlorine gas for 16h. Surface sterilized seeds were germinated on ¼ Gamborg's medium. Hairy roots were generated by cotyledonary node method (Kandoth et al., 2011). Briefly, cotyledons from 7-9 day old seedlings were cut and infected with *A. rhizogenes* K599 carrying the sgRNA construct by incubating for 2 days. Cotyledons were washed for 1-2 hours on a rotator in ¼ Gamborg's liquid with 238ug/ml Timentin to inhibit *A. rhizogenes* growth. Cotyledons were removed and transferred with the cut surface facing up onto plates with MXB medium (MS basal nutrient salts (Caisson Laboratories, North Logan, UT), 1× Gamborg's B-5 vitamins, 3% sucrose, 0.7% Phyto Agar (Research Product International Corp), pH 5.7) with 238ug/ml Timentin. After 10-14 days, hairy roots formed were cut and propagated on fresh MXB plates with Timentin. Transgenic hairy roots were selected using green fluorescence marker under the microscope. Root tips were subcultured three times and then used for nematode infection experiments.

Genomic DNA extraction from hairy roots

During hairy roots propagation, approximately 1 to 1.5cm of the top portions of each hairy root was collected in 1.5ml eppendorf tubes aseptically. Samples were

immediately frozen into liquid nitrogen and stored at -80°C until DNA extraction.

REExtract-N-Amp Plant PCR kits (Sigma) were used for genomic DNA extraction. For DNA extraction, each root sample was smashed with sterilized pestles. In each tube, 100 µL of the extraction solution was added. After a brief vortex mixing, the tubes were incubated in water bath at 95°C for 10 minutes. Following this, 100ul of the dilution solution was added to each tube and vortexed. The DNA extracts were stored for a brief period at 4°C, if not used immediately for PCR.

Sequencing screening for modifications in *GmSHMT08*

Genomic DNA extracted from soybean hairy roots were PCR amplified with specific primers flanking the sgRNA target sites (Table 2.1). PCR products were gel purified and sequenced by MU DNA core facility to determine modifications.

Target	PCR primers	Sequence 5'-3'
EXF67/63 <i>GmSHMT08</i>	GmA2Shmt_tilling _F	ACAACACTCTCTCTTCTCGC
	SHMT08 crispr seq R1	CTGCAGCGAAGGGAACAC
EXF67/63 <i>GmSHMT05</i>	SHMT crisp 5F	CCAATCCTGGGGCGTCAATGTG
	CRISPSHMT5-seq R	AGCGCCACGGCAAGAGCC

Target	Sequencing primer	Sequence 5'-3'
EXF67/63 <i>GmSHMT08</i>	Forrest SHMT sac1 F	AATTGAGCTCATCGCCTCCGAGA
EXF67/63 <i>GmSHMT05</i>	SHMT crisp 5F	CCAATCCTGGGGCGTCAATGTG

Off-targeting analysis

GFP positive transgenic hairy roots of soybean that were edited by p201GCas9-*GmSHMT08* were PCR amplified with *GmSHMT05* specific primers flanking the

homologous region. PCR products were gel purified and sequenced to look for any modifications through MU DNA core facility.

Off-targeting gene: EXF67/63 <i>GmSHMT05</i>		Sequence 5'-3'
PCR primers	SHMT crisp 5F	CCAATCCTGGGGCGTCAATGTG
	CRISPSHMT5-seq R	AGCGCCACGGCAAGAGCC
Sequencing primer	Forrest SHMT sac1 F	AATTGAG CTCATCGCCTCCGAGA

PCR screening for deletions in *GmSHMT08*

Genomic DNA extracted from transgenic soybean hairy roots were amplified using primers flanking the region spanned by both sgRNAs. Electrophoresis of PCR products were done to compare PCR product size from the transgenic hairy roots to that of vector transformed control (unmodified) hairy roots.

Target	PCR primer	Sequence 5'-3'
EXF67 <i>GmSHMT08</i>	CRISPR SHMT8_seq F1C	TGGGGGCTAGATCTCCG
	SHMT8 crispr seq R1	CTGCAGCGAAGGGAACAC
EXF63 <i>GmSHMT08</i>	CRISPRessexSHMT8-seqF1	TGGGGCTAGATCTCCCC
	SHMT8 crispr seq R1	CTGCAGCGAAGGGAACAC

Nematode infection assay

Transgenic hairy roots that were propagated for 3-4 times on MXB plates and freed from *A. rizogenus* were used for nematode infection. 15-20 healthy independent hairy root lines were used for each construct. Vector transformed transgenic hairy roots of ExF67 (resistant) and ExF63 (susceptible) were used as controls. Each plate contained two 4-5 cm long hairy roots. PA3 SCN juveniles were hatched at 27°C chamber in a two-

day incubation period. J2 pre-parasitic juveniles were surface sterilized with 1ml of 0.004% [w/v] mercuric chloride, 0.004% [w/v] sodium azide, and 0.002% [v/v] Triton X-100 for 8 min followed with 3-4 times of washing with sterilized water and finally resuspended in 0.1% sterilized agarose. Each hairy root was inoculated with a total of 450 J2s in a volume of 30-35 μ l. Roots were inoculated approximately 1 cm above the root tip and 1 cm below the root top. The plates were placed in the dark at room temperature after sealing with two layers of parafilm. After 30 days, cysts were counted under the microscope. The number of cysts were plotted and analyzed for statistical significance using t-test using GraphPad PRISM software.

Results

Application of CRISPR/Cas9 to knock out *GmSHMT08* in soybean hairy roots

In previous studies, several reverse genetic approaches were used to silence the *GmSHMT08* gene including VIGS, RNAi, and the identification of TILLING mutants (Liu et al., 2012). Therefore, *GmSHMT08* was used here as a target in a proof-of-concept study to validate the gene-knock-out efficiency of the CRISPR/Cas9 system in SCN resistant soybean hairy roots.

Two different types of CRISPR/Cas9 constructs were designed, constructed, and tested for gene modification efficiency of *GmSHMT08* (Figure 2.1). The first type of CRISPR/Cas9 construct was designed to include one gRNA and was based on the report by Jacobs et al (2015). Five gRNAs were designed using online sources (<http://www.genome.arizona.edu/crispr/>, <http://cbi.hzau.edu.cn/cgi-bin/CRISPR>) to target the Forrest *GmSHMT08* cDNA sequence (Figure 2.2). The gRNA #2 could not be cloned into the vector and therefore was not pursued further. The remaining four gRNAs were cloned into the p201GCas9 vector backbone to generate constructs SHMT8-1, SHMT8-3, SHMT8-4, and SHMT8-5. Each construct was transformed into the SCN-resistant soybean EXF67 RIL background to generate hairy roots to test for modification efficiency (Table 2.1). Modification of *GmSHMT08* was assessed by amplifying the full-length cDNA using PCR and then sequencing the gene fragments. Hairy roots represent a chimera of different target site modifications, so sequence chromatographs were analyzed for polymorphic sequence at the target site (Figure 2.3). No modifications were detected in hairy roots transformed with SHMT8-3 and SHMT8-5 and were not pursued further. Five biological replicates of hairy root transformation were conducted for SHMT8-1 and

SHMT8-4 and anywhere from 6-26 independent hairy root lines per replicate were tested for modification. In total, 68 and 89 independent hairy root lines were sequenced for constructs SHMT8-1 and SHMT8-4, respectively (Table 2.1). Of these, 48.5% (SHMT8-1) and 37.1% (SHMT8-4) of the hairy root lines tested positive for *GmSHMT08* modification (Table 2.1). Off target modification of *GmSHMT05*, a closely related copy of *GmSHMT08*, located on chromosome 5 was also investigated. Sequencing of 20 hairy root lines did not identify any off-target modifications of *GmSHMT05*.

The second type of CRISPR/Cas9 construct utilized two gRNAs to induce a specific 130 bp nucleotide deletion of the gene target and was designed based on the reports of Mao et al., (2013) and Feng et al. (2013). Because SHMT8-1 and SHMT8-4 were found to be effective in specifically modifying *GmSHMT08* in soybean hairy roots, the sequences of these two gRNAs were used to generate pCam-GWOX-Cas9-SHMT8 (named as SHMT8-14) using the pCam-GWOX-Cas9 vector backbone (Figure 2.3; Table 2.1). The construct was transformed into the SCN-resistant soybean EXF67 RIL line and the SCN-susceptible soybean EXF63 RIL line to generate hairy roots to test for modification efficiency. In order to detect modifications, primers were designed to amplify a 500 bp fragment of *GmSHMT08* spanning the target site using PCR. PCR products were resolved by gel electrophoresis to identify hairy roots harboring a deletion in *GmSHMT08* (Figure 2.5). Nearly 70% of the hairy root lines analyzed in each genetic background contained a deletion (Figure 2.6). Several roots where a deletion could not be resolved by gel electrophoresis were sequenced. In many cases, small deletions were also present in these lines (Table 2.2).

Taken together, these results indicate that CRISPR/Cas9 is effective in knocking out genes in soybean using a hairy root transformation approach.

Evaluation of CRISPR/Cas9 *GmSHMT08* soybean hairy roots in a nematode infection assay

Two single gRNA constructs (SHMT8-1 and SHMT8-4) and one targeted deletion construct (SHMT8-14) proved to be effective in editing *GmSHMT08* in soybean hairy roots. We then wanted to test whether the CRISPR/CAS9 protocol could be applied to our hairy root infection assay pipeline to extend its use for the functional analysis of soybean genes involved in resistance to SCN. Therefore, we next measured nematode development on hairy roots generated with each of the three CRISPR/CAS9 constructs or the vector control in the SCN resistant EXF67 RIL background harboring the Forrest type *GmSHMT08*.

Hairy roots were excised from EXF67 soybean cotyledons and propagated two to three times prior to inoculation with SCN infective second-stage juveniles. The numbers of adult female nematodes (cysts) were counted at 30 days post-inoculation (dpi). A total of 15 to 25 independent hairy roots lines were generated for SHMT8-1, SHMT8-4, and an empty p201GCas9 vector in each experiment. Hairy root lines generated using SHMT8-1 showed a statistically significant increase in nematode infection compared to empty vector control roots in three biological replicates of the experiment (Figure 2.7). Hairy root lines generated using SHMT8-4 showed a statistically significant increase in nematode infection compared to empty vector control roots in two of the three biological replicates of the experiment (Figure 2.7). For comparison, the average nematode infection

level in SHMT8-14 EXF67 lines was similar to that of the SCN-susceptible control line EXF63 in one biological replicate of the experiment.

Similarly, SHMT8-14 EXF67 hairy root lines showed a statistically significant increase in nematode infection compared to vector control roots in four biological replicates of the experiment (Figure 2.8). The average nematode infection level in SHMT8-14 EXF67 lines was similar to that of the SCN-susceptible control line EXF63 in two biological replicates of the experiment (Figure 2.8 C). Taken together, these data confirm our prior VIGS, RNAi, and TILLING results indicating that disruption of the Forrest type *GmSHMT08* gene function increases SCN susceptibility in SCN resistant soybeans. Moreover, we validate CRISPR/Cas9 as an efficient method to test soybean genes for a role in resistance to SCN using our transgenic hairy root pipeline.

Using CRISPR/Cas9 to test *GmSHMT08* for a role in soybean basal resistance

Two amino acid polymorphisms distinguish the SCN-resistant soybean cv. Forrest type *GmSHMT08* and the SCN-susceptible soybean cv. Essex (and Williams 82) type *GmSHMT08*. We hypothesized that *GmSHMT08* likely plays a critical function in a compatible plant-nematode interaction due to a greater demand in folate metabolism for establishment of the nematode feeding cell. However, prior experiments in the lab using VIGS and RNAi to silence *GmSHMT08* in the SCN-susceptible soybean cv. Essex (and Williams 82) resulted in an unexpected hypersusceptibility to the nematode suggesting that this gene may be functioning in soybean basal resistance to SCN (P. Kandoth & M.G. Mitchum, unpublished). Due to the potential cross-silencing of closely related soybean *GmSHMT* sequences in VIGS and RNAi experiments, we could not attribute this

phenotype solely to the Essex-type *GmSHMT08*. To test this further, we transformed the SHMT8-14 CRISPR/Cas9 construct previously shown to be specific to *GmSHMT08* into the SCN susceptible soybean EXF63 RIL line to test this gene for a role in basal resistance to the nematode in soybean. In four biological replicates of the experiment, only one replicate showed a statistically significant increase in susceptibility of the SHMT8-14 EXF63 lines compared to empty vector control roots (Figure 2.9). Hairy root infection assays suffer from poor infection rates leading to a high degree of root to root variability in cyst counts within a single genotype. This is why detection of subtle changes in susceptibility can be masked unless large numbers of roots are evaluated. However, root propagation is very labor intensive and limits the number of roots that can be included in the assay. Interestingly, in the infection assay where a statistically significant increase in susceptibility was observed, the average number of cysts in both susceptible vector control and SHMT8-14 lines was 50% higher than the other three replicates indicative of a good infection rate. Taken together, we were unable to confirm that disruption of the Essex type *GmSHMT08* increases susceptibility of soybean to SCN. Additional methods of testing such as using the soybean composite plant system (Guo et al., 2011) and/or stable whole plant soybean transformation will be required to further evaluate a role for *GmSHMT08* in basal resistance.

Analysis of *GmSHMT05* in soybean

The soybean genome sequence (SCN susceptible cv. Williams 82) contains a *SHMT* gene that is closely related to *GmSHMT08* (Glyma08g11490.2) on chromosome 5 called *GmSHMT05* (Glyma05g28490.1) (Schmutz et al., 2010). TILLING mutants identified in

our lab have shown that *GmSHMT05* in the SCN resistant cv. Forrest does not contribute to SCN resistance (Liu et al., 2012; P. Kandoth & M.G. Mitchum, unpubl.). A recent phylogenetic study of SHMTs in soybean has shown that *GmSHMT08* and *GmSHMT05* have the same length and gene structure which could be the result of a whole genome duplication event that happened 13 million years ago (Wu et al., 2016). Although *GmSHMT05* may not play a role in SCN resistance, we hypothesized that it may play a critical role in soybean root growth and development especially considering that the Peking-type of resistance to SCN (i.e. cv. Forrest) has been associated with yield drag in the field in the absence of SCN pressure (Mitchum, 2016). Thus, we wondered if this might be attributed in part to the gain of function of the *GmSHMT08* in SCN resistance and the inability of *GmSHMT05* to fully complement the required level of functional (i.e., normal) SHMT necessary for plant growth and development in certain conditions in an SCN resistant background. Thus, knocking out *GmSHMT05* in the SCN resistant background might impact root growth and development, if *GmSHMT08* gained a new and unique function in SCN resistance. We set out to test this using CRISPR/Cas9.

We first cloned and sequenced full-length *GmSHMT05* cDNA sequences from Forrest, Essex, EXF63, and EXF67 to aid in the design of gRNAs targeting *GmSHMT05*. Upstream and downstream flanking primers were designed based on the Williams 82 genome sequence (soykb.com) according to the predicted start and stop of *GmSHMT05*. Amplified cDNA products generated from total RNA isolated from leaf tissues of each genotype were sequenced. Multiple nucleotide sequence alignment indicated that *GmSHMT05* coding region sequences were the same in EXF67, EXF63, Essex and Forrest. A nucleotide alignment between *GmSHMT05* and the Forrest type *GmSHMT08*

indicated that these two sequences were 95.8% identical (Figure A.1). Seven amino acid polymorphisms were identified between GmSHMT05 and the Forrest type GmSHMT08 (Figure 2.10, Table 2.3). A nucleotide alignment between *GmSHMT05* and the Essex type *GmSHMT08* indicated that these two sequences were 95.9% identical (Figure A.2). Five amino acid polymorphisms were identified between *GmSHMT05* and the Essex type *GmSHMT08* (Figure 2.10). The two amino acid polymorphisms accounting for SCN resistance in Forrest type *GmSHMT08* (Liu et al., 2012) were absent from *GmSHMT05*, which is consistent with our inability to detect a role for this SHMT in resistance to SCN. However, the five amino acid polymorphisms between GmSHMT05 and Essex type GmSHMT08 may reflect functional differences that are not completely redundant during plant growth and development.

Two CRISPR/Cas9 constructs provided by our collaborators at University of Georgia (W. Parrott and T. Jacobs) were designed to either target *GmSHMT05* alone (Table 2.4) or both *GmSHMT08* and *GmSHMT05* using the single gRNA CRISPR/Cas9 system (Jacobs et al., 2015). Each construct was transformed into hairy roots and modification efficiency was calculated by sequencing *GmSHMT05* in individual roots. No obvious hairy root growth phenotypes were noted when SHMT5 was transformed into either EXF63 or EXF67; however, hairy roots generated from EXF67 soybeans had severe bacteria contamination in the first two replicates. Consistent with earlier TILLING results indicating that *GmSHMT05* does not contribute to SCN resistance, targeting *GmSHMT05* by CRISPR/Cas9 in the SCN resistant background (EXF67) did not alter the resistant phenotype in two biological replicates (Figure 2.11 A-B). In addition, targeting *GmSHMT05* in the SCN susceptible background (EXF63) did not alter susceptibility to

the nematode (Figure 2.11 C-D). The average modification rate of SHMT5 in EXF67 was 48% (n=33) compared with 77% in EXF63 (n=22). The SHMT58 construct was also used to generate hairy roots in EXF67 soybeans, but unfortunately no *GmSHMT05* or *GmSHMT08* modifications were detected (Table 2.4).

Discussion

Here we demonstrated that two different types of CRISPR/Cas9 systems published previously (Jacobs et al., 2015; Feng et al, 2013, Mao et al, 2013), utilizing either a single gRNA or two gRNAs, to target *GmSHMT08* in SCN resistant soybeans were effective in knocking out *GmSHMT08*. However, modification efficiency and ease of detection varied between the two methods. Four different gRNAs were designed to target *GmSHMT08*. Selection of gRNAs was based on multiple rules such as being adjacent to a PAM site, starting with a “G” and were ranked by online tools. However, efficiency of modification cannot be predicted simply according to gRNA sequences. Each gRNA was cloned independently into the p201GCas9 vector which utilized the *Medicago truncatula* U6.6 promoter and human codon-optimized Cas9. The gRNA constructs were then used to transform the SCN-resistant genotype to generate soybean hairy roots. Only two (SHMT8-1, SHMT8-4) out of the four constructs showed modification to *GmSHMT08*. In addition, the modification rate for each of the two constructs did not exceed 50%. Sanger sequencing of an amplified fragment spanning the target site was required to detect modifications and represented a mixture of modifications at the gRNA target region. These results are consistent with previous reports indicating that hairy roots represent a chimera of somatic modification events (Jacobs et al., 2015). Chimeric tissues were evident based on the detection of a mixture of different types of mutations such as deletions, insertions and combined mutations (Pan et al., 2016). The CRISPR/Cas9 system starts to perform DNA cleavage when all of the CRISPR/Cas9 components are transformed into a soybean cell. As cells divide, different

target gene modifications can occur resulting in daughter cells with different genotypes. Therefore, the parenchyma cells on the callus formed on soybean cotyledons could be chimeric tissues. Consequently, each hairy root may generate different types of mutations in the target gene as it continues to grow. Analysis of chimeric tissue complicates the analysis of modification efficiency in hairy roots. The reasons for the lack of detectable modifications with two of the gRNAs are unknown. The less than 50% modification rate observed for gRNAs 8-1 and 8-4 may be due to limited sensitivity of detecting modifications by sequencing if modifications were introduced late in hairy root formation, the choice of gRNA sequences, or the promoter sequences driving the expression of the gRNA and the Cas9.

The dual gRNA CRISPR/Cas9 construct (SHMT8-14) showed a significantly higher percentage of soybean hairy roots with targeted deletions in *GmSHMT08*. The gRNA-1 and gRNA-4 were cloned into the 35S-Cas9-SK system which utilized *Arabidopsis* AtU6-26 promoter and human codon-optimized *Streptococcus pyogenes* Cas9 (hspCas9) (Cong et al, 2013). Amplification of the *GmSHMT08* gene sequence spanning the targeted deletion site from hairy roots by PCR and separation by gel electrophoresis demonstrated a high frequency of deletions in the target gene but the size of the nucleotide deletion varied. Further sequencing of the amplified fragments that did not show a detectable shift by gel electrophoresis identified deletions as small as 2 bp. Using both approaches, deletions ranging from 2 bp to 300 bp were detected. The average deletion rate as determined based on gel shift from three biological replicates in both SCN resistant and susceptible backgrounds was approximately 67%, which is 20% higher than the single gRNA CRISPR/Cas9 system. However, gene function can be disrupted by

a single nucleotide deletion. PCR amplification and gel resolution to detect modified roots is a useful screening tool, but is not a reliable predictor of the mutation rate as many roots did not show a gel shift, but harbored small deletions. When using a combination of gel shift and sequencing, a mutation rate of >90% was detected with the dual gRNA CRISPR/Cas9 construct SHMT8-14. The higher percentage of gene modifications using the dual gRNA construct could be due to a higher level of expression of the AtU6.26 promoter relative to the MtU6.6 promoter in soybean hairy roots. A recent study demonstrated that the native soybean GmU6-16g-1 promoter cloned from cv. Williams 82 showed more efficient gene modification in soybean hairy roots than the Arabidopsis AtU6-26 promoter, so this could be improved further (Du et al, 2016).

Hairy roots generated using both CRISPR/Cas9 systems targeting *GmSHMT08* were subjected to nematode infection assays. All roots for each construct were included in the infection assay, not just those confirmed to carry modifications. In each of three biological replicates with constructs 8-1 and 8-4 and four biological replicates with construct 8-14 targeting *GmSHMT08* in the resistant background, all but one replicate with 8-4 (Figure 2.7A) showed a statistically significant increase in susceptibility to SCN. In both approaches, it was not necessary to pre-select hairy roots with confirmed modifications of the target gene for the infection assays. However, pre-selection would be advised when testing new genes for a role in SCN resistance, as well as testing for altered susceptibility to SCN by knocking out genes in a susceptible genotype. The dual gRNA deletion approach offers the advantage of being able to quickly confirm modification by PCR and gel shift, allowing rapid and stringent selection of hairy roots for inclusion in infection assays.

The SCN infection assay results of *GmSHMT08*-modified/deleted hairy root lines indicated that knocking out *GmSHMT08* in SCN resistant soybeans could significantly increase susceptibility to SCN consistent with our previously published data (Liu et al., 2012). The cell death response of the syncytium in resistant soybeans may be related to the crosstalk between nematode secreted proteins and host defense gene expression (Kyndt et al., 2013). The resistant Forrest type *GmSHMT08* at the *Rhg4* locus encodes the ubiquitous enzyme serine hydroxymethyltransferase, which is essential in one-carbon folate metabolism (Liu et al., 2012). The two amino acid polymorphisms at *GmSHMT08* between resistant and susceptible cultivars may alter the enzymatic activity leading to perturbations in folate homeostasis during syncytium formation. The failure of syncytium establishment may be related to the metabolic stress associated with folate deficiency. Folate is also a crucial nutrient for SCN to complete its life cycle. SCN does not produce folate by itself, so it must rely on host plants to absorb folate. Therefore, a folate deficiency in resistant soybeans might lead to the demise of SCN and the lack of molecular stimuli to maintain the syncytium (Goverse et al., 2014). Deciphering the underlying mechanism of SHMT resistance is a current active area of research in the lab.

Our work targeting *GmSHMT08* in the SCN-resistant soybean background served as a “proof of concept” to demonstrate the utility of CRISPR/Cas9 in our soybean hairy root pipeline for studying the role of soybean genes in resistance to SCN. We then focused on application of CRISPR/Cas9 to knock out *GmSHMT08* in the SCN-susceptible EXF63 RIL, which carries the Essex type *GmSHMT08*. Unpublished results in the lab using VIGS and RNAi to silence *GmSHMT08* in SCN-susceptible EXF63 and W82 showed an unexpected result of increased susceptibility to SCN. We had

hypothesized that developing syncytia likely require a high demand for 1-C metabolism and any perturbations to folate homeostasis by silencing *GmSHMT08* in susceptible soybeans would compromise syncytium formation and thereby decrease susceptibility to SCN. However, the observed increased susceptibility to the nematode in *SHMT* silencing experiments suggests that this enzyme may be playing a role in basal resistance to SCN, also known as pathogen- or microbial-associated molecular pattern (PAMP or MAMP) triggered immunity or PTI (Jones et al., 2006, Dodd et al., 2010). Basal resistance is the first line of defense to protect plants against pathogens (Muthamilarasan et al., 2013). Pattern recognition receptors (PRRs) which are localized at the cell surface could recognize PAMP or MAMPs to activate a defense signaling cascade. Knocking out plant genes involved in PTI typically results in increased susceptibility to the pathogens (Huot & Yao, 2014). The intracellular migration pattern of SCN could activate plant-derived elicitors of basal resistance or the plant may detect surface molecules or secreted effectors that elicit PTI. In susceptible plants, basal resistance has been shown to be suppressed by the nematode as they migrate towards the vascular cylinder to establish a feeding site within the roots (Holbein et al., 2016).

Due to the potential cross-reactivity of VIGS and RNAi silencing constructs (both utilize a 300bp *GmSHMT08* sequence) with other soybean *GmSHMTs*, we have not been able to attribute this increased susceptibility phenotype to silencing of *GmSHMT08* alone. The SHMT8-14 CRISPR/Cas9 dual gRNA construct is highly specific for *GmSHMT08*. This construct was shown not to modify the closely related *GmSHMT* gene on chromosome 5 (*GmSHMT05*). SHMT8-14 was transformed into the SCN-susceptible EXF63 RIL line to generate modified roots for infection assays. In four biological

replicates of the experiment, only one replicate showed a statistically significant increase in susceptibility of the SHMT8-14 EXF63 lines compared to empty vector control roots. Hairy root infection assays suffer from poor infection rates leading to a high degree of root to root variability in cyst counts within a single genotype. This is why some of the roots, despite being confirmed for modification, do not exhibit an infection phenotype. Detection of subtle changes in susceptibility can be masked unless large numbers of roots are evaluated. However, root propagation is very labor intensive and limits the number of roots that can be included in the assay. Interestingly, in the infection assay where a statistically significant increase in susceptibility was observed, the average number of cysts in both susceptible vector control and SHMT8-14 lines was 50% higher than the other three replicates indicative of a good infection rate. Taken together, we were unable to conclude that disruption of the Essex type *GmSHMT08* increases susceptibility of soybean to SCN using this assay. Additional methods of testing such as the soybean composite plant system (Guo et al., 2011) and/or stable whole plant soybean transformation will be required to further evaluate a role for *GmSHMT08* in basal resistance.

GmSHMT05 (Glyma05g28490.1) on chromosome 5 is closely related to *GmSHMT08* (Schmutz et al., 2010), but does not play a role in SCN resistance (Liu et al., 2012; P. Kandoth & M.G. Mitchum, unpubl.). Although *GmSHMT05* may not play a role in SCN resistance, we hypothesized that it may play a critical role in soybean root growth and development especially considering that the Peking-type of resistance to SCN (i.e. cv. Forrest) has been associated with yield drag in the field in the absence of SCN pressure (Mitchum, 2016). Thus, we wondered if this might be attributed in part to the gain of

function of the *GmSHMT08* in SCN resistance and the inability of *GmSHMT05* to fully complement the required level of functional (i.e., normal) SHMT necessary for plant growth and development in certain conditions in an SCN resistant background. Here we used CRISPR/Cas9 to knock out *GmSHMT05* in the SCN resistant background. However, despite confirmed modifications to *GmSHMT05*, no notable impact on root growth and development was observed. There are 14 *GmSHMTs* in soybean (Wu et al., 2016), so functional redundancy of SHMTs in plant growth and development is likely.

Table 2. 1. Summary of two types of CRISPR/Cas9 constructs made targeting on *GmSHMT08*.

Type of gRNA	Name	Vector	gRNA promoter	Sequence 5'-3'	Selection	Target position on <i>GmSHMT08</i>	Test genotype	Modification of <i>GmSHMT08</i>					Number of total amplified events	Avg mod rate from sequencing
								rep1	rep2	rep3	rep4	rep5		
Single gRNA	SHMT8-1	p201GC9	MtU6.6	AGGGAAGTCGCTGATAAGTG	GFP	604-624	EXF67	8/11	2/6	3/9	10/22	10/20	68	48.50%
	SHMT8-3	p201GC9	MtU6.6	AGGCCAAAACATCAATACTG	GFP	541-561	EXF67	0/10	-	-	-	-	10	0%
	SHMT8-4	p201GC9	MtU6.6	GCTTGGGGGGCCACGTGCG	GFP	735-755	EXF67	7/17	3/9	4/13	7/26	12/24	89	37.10%
	SHMT8-5	p201GC9	MtU6.6	GTCAACGTCCAGCCCTACTC	GFP	235-255	EXF67	0/14	-	-	-	-	14	0%
Dual gRNA	SHMT8-14	pCam-GWOX-Cas9	AtU6.26	AGGGAAGTCGCTGATAAGTG	GFP	604-624 and 735-755	EXF67	Nucleotide deletion of <i>GmSHMT08</i>					66	60.60%
				rep1				rep2	rep3	rep4	rep5			
				GCTTGGGGGGCCACGTGCG			EXF63	15/23	15/23	15/22	10/21	12/18	61	67%

Table 2. 2. Nucleotide deletion analysis. Examples of EXF67 SHMT8-14 hairy root sequencing compared with PCR/Gel estimation on the deletion of *GmSHMT08*.

Hairy root number	Estimated fragment size	Sequencing details
1	~300bp	Complete deletion of 150bp in between two targets
2	~450bp	Modification of 186bp including two targets
3	500bp	Sequence is a mixture of "N"s between two targets
4	500bp	4bp missing in gRNA-4
5	~480bp	Missing 2bp in gRNA-1 and missing 4bp in gRNA-4
6	~350bp	Complete deletion of 130bp between gRNA-1 and gRNA-4
7	~300bp	220bp deletion from upstream of gRNA-1 till a little bit downstream of gRNA-1. gRNA-1 is deleted.

Table 2. 3. Nucleotide and Amino acid comparison between Forrest type and Essex type *GmSHMT08*

	Nucleotide (Amino acid)	
	389 (130)	1072 (358)
Essex-SHMT8	C (P)	A (N)
Forrest-SHMT8	G (R)	T (Y)

Table 2. 4. Summary of 201GCas9 constructs modification efficiency on *GmSHMT05*. Note: “cont” means severe bacterial contamination and limited numbers of roots could be collected. nt = not tested

Name	Vector	gRNA promoter	Sequence 5'-3'	Selection	Target position (5'-3')	Test genotype	Modified events/total amplified events				Number of total amplified events	Avg mod rate from sequencing	Off-targeting percentage
							rep1	rep2	rep3	rep4			
SHMT5	p201GC9	MtU6.6	GGAAATTGCGGATAAGTGCG	GFP	606-626 on <i>GmSHMT05</i>	EXF63	7/10	10/12	-	-	22	77%	nt
						EXF67	cont	cont	7/11	8/18	33	48%	0%
SHMT58	p201GC9	MtU6.6	GGGCGCACACTAGCGGCCTTG	GFP	648-668 on <i>GmSHMT05</i> and <i>GmSHMT08</i>	EXF67	cont	0/12	-	12	0%	nt	

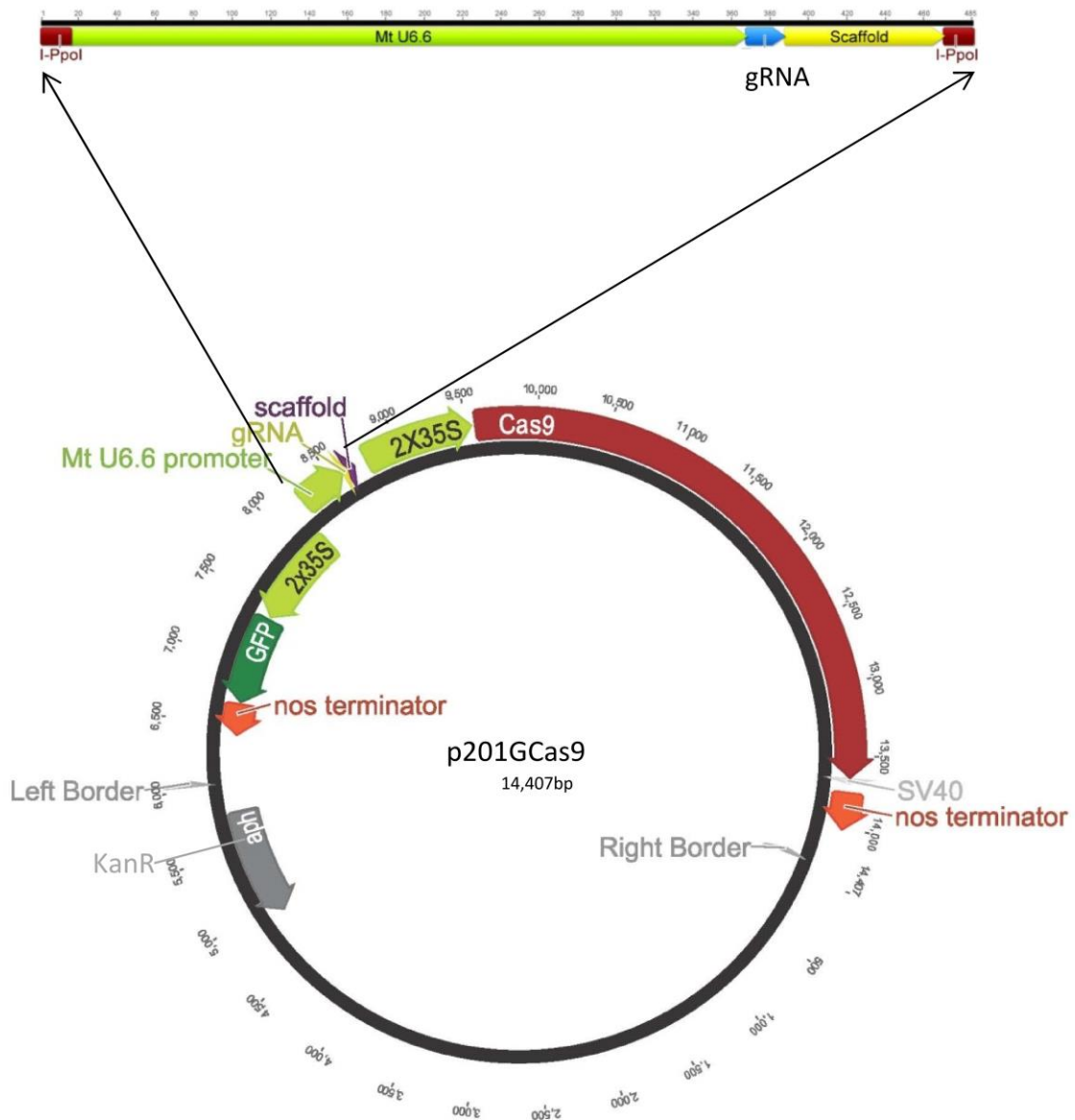


Figure 2. 1. Vector used for single gRNA construct. The plasmid p201GCas9 was the binary vector used for hairy root transformations. In this construct, Cas9 is driven by double 2X35S and GFP and was used to select transformed roots. I-PpoI sites accept the gRNA insertion from pUC gRNA shuttle and the *Medicago truncatula* U6.6 promoter drives the gRNA.

```

>cDNA of GmSHMT08 (1416bp)
  1 atggatccag taagcgtgtg gggtaacacg cccttggcga cggtaggatcc cgagatccat
  61 gacctcatcg agaaggagaa gcgccgtcaa tgcccgaggaa tcgagctcat cgcctccgag →
 121 aacttcacct ccttcgccgt catcgaggcc ctccggcagcg ctctcacgaa caaatactcc
 181 gagggatcgc cgggcaaccg ctactacggc ggcaatgaat acatcgacca gatcgaaaac
 241 ctctgccgct cagcgcacct ccaagccttc cacctcgacg cccaatcctg gggcgtcaac
 301 gtccagecct actccggctc cccggccaac ttccgccgct acaccgccgt cctcaacccc
 361 cacgaccgca tcatggggct agatctccgc tccggcggcc acctcaccca cggctactac
 421 acctccggcg gaaagaagat ctccgccacc tccatthact tcgagagtct cccttacaag
 481 gtaaaactca ccaccggcta catcgactac gaccgcttgg aagaaaaagc cctagacttc
 541 aggccaaaac tcataatctg cggtggcagc gcgtaccctc gcgattggga ctacaaacgt
 601 ttcagggaag tcgctgataa gtgcggagca ttgcttctct gcgacatggc gcacactagc
 661 ggccttgtgg ccgcgagga agtgaacagc cccttcgagt attgcgacat tgtgaccacc
 721 acgactcaca agagcttgcg gggcccacgt gcggggatga tcttttaccg gaagggcccc
 781 aagccgccga agaaggggca gccggagaac gcgggtttatg atttcgagga caagattaac
 841 ttccggtgtg tcccttcgct gcagggtggg ccccaacaacc accagatcgg tgctctcgcc
 901 gtggcgctga agcaggccgc gtcgccggg ttttaaggcct acgcgaagca ggttaaggcg
 961 aacgccgctg cgcttgaaa atacttgatg gggaaagggt acagccttgt cactggcggg
1021 acggagaacc atcttgttt gtgggatctg agacctcttg gattgactgg gtataagggtg
1081 gagaaactct gtgatctctg taacattact gttaacaaga acgctgtttt tgggtgatagc
1141 agtgccttgg cccctgggtg agtgcgaatt ggtgcccctg ccatgacttc taggggtttg
1201 gttgaaaaag actttgagca gattggtgag ttccttcacc gtgctgtgac tctcacactg
1261 gagatccaga aggagcatgg caaacttctc aaggatttca acaagggtct cgtcaacaac
1321 aaggctattg aagatctcaa agctgatggt gagaagttct ctgccttggt tgacatgcct
1381 ggcttctctg tatctgaaat gaagtacaag gattag

```

Figure 2. 2. Positions of gRNAs on Forrest type *GmSHMT08*. Four different gRNAs were designed to target *GmSHMT08* and are highlighted as gRNA-1 (yellow), gRNA-3 (grey), gRNA-4 (purple), and gRNA-5 (green). PAM sites are shown in blue color. Black arrows represent two sequencing primers: Forrest SHMT sacI F (forward) and SHMT08 crispr seq R1 (reverse) used for amplifying *GmSHMT08* in hairy roots gDNA to check mutations.

A

Hairy root lines	Sequence alignment with Forrest <i>GmSHMT08</i>
Forrest	tttc aggg aagtcgctgataaagtc ggg agcattgcttctctgcgacatggcgcacactag
EF67SHMT8-1-14-2-ud	TTTCAGGGAAAGTCGCTGATAANGGNGGNNTTNGNTTCTCNGNNANNNGNNNCANANN-N
EF67SHMT8-1-2-ud	TTTCAGGGAAAGTCGCTGATAAGTGC GGAGCATTGCTTCTCTGCGACATGGCGCACACTAG
EF67SHMT8-1-17-1-ud	TTTCAGGGAAAGTCGCTGATAANGGCGNANNNTTGCNCCNN-TGNGNNNNNNNNNNNNNN
EF67SHMT8-1-17-2-ud	TTTCAGGGAAAGTCGCTGATAAGTGC GGAGCATTGCTTCTCTGCGACATGGCGCACACTAG
EF67SHMT8-1-29-ud	TTTCANNGGAAGTCGCTGATNNNNNGNNGANNNTGCTTCTCCTGCGNTGNGNCGCACANNAN
EF67SHMT8-1-14-3-ud	NNTC TGGNNNNNGCTANNNNNG CANANTAGNNNCNNTGNGNNNN--NNNNNNNNNN
EF67SHMT8-1-38-1-ud	TTTCAGGGAAAGTCGCTGATAAGTGC GGAGCATTGCTTCTCTGCGACATGGCGCACACTAG
EF67SHMT8-1-19-ud	TTTCAGGGAAAGTCGCTGATAANGGNGNANNATNGNTTCTCTGNNNNANGNNGCACACT-N
EF67SHMT8-1-24-ud	TTTCAGGGAAAGTCGCTGATAAGTGC GGAGCATTGCTTCTCTGCGACATGGCGCACACTAG
EF67SHMT8-1-41-2-ud	TTTCAGGGAAAGTCGCTGATAANNNGGGAGCATTGCTTCTCTGCGACATGGCGCACACTAG
EF67SHMT8-1-15-ud	TTTCAGGGAAAGTCGCTGANNNGGNNNNNNTNNNTTCTCNGNNNNNGGNNCANNNNNN
EF67SHMT8-1-38-ud	TTTCAGGGAAAGTCGCTGATAAGTGC GGAGCATTGCTTCTCTGCGACATGGCGCACACTAG
EF67SHMT8-1-17-3-ud	TTTCAGGGAAAGTCGCTGATAAGTGC GGAGCATTGCTTCTCTGCGACATGGCGCACACTAG
EF67SHMT8-1-41-1-ud	TTTCAGGGAAAGTCGCTGATAAGTGC GGAGCATTGCTTCTCTGCGACATGGCGCACACTAG
EF67SHMT8-1-21-3-ud	TTTCAGGGAAAGTCGCTGATAAGTGC GGAGCATTGCTTCTCTGCGACATGGCGCACACTAG
EF67SHMT8-1-9-1-ud	TTTCAGGGAAAGTCGCTGATANNNGNNGANNNTTGNNTTCTCNGNNANNNGGCGCACACTAN
EF67SHMT8-1-1-ud	TTTCAGGGAAAGTCGCTG---AGTGC GGAGCATTGCTTCTCTGCGACATGGCGCACACTAG
EF67SHMT8-1-3-1-ud	TTTCAGGGAAAGTCGCTGATAANGNNGNANNNTNNNTTCTCNGNNNNNGGNNNCANNNN
EF67SHMT8-1-7-ud	TTTCAGGGAAAGTCGCTGATAAGTGC GGAGCATTGCTTCTCTGCGACATGGCGCACACTAG
EF67SHMT8-1-21-1-ud	TTTCAGGGAAAGTCGCTGATAAGTGC GGAGCATTGCTTCTCTGCGACATGGCGCACACTAG

B

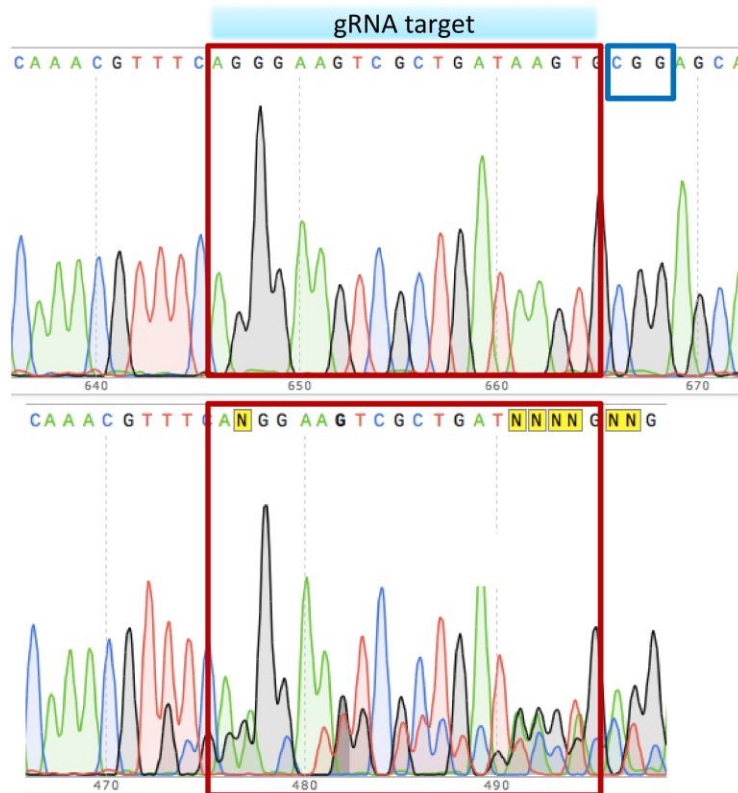


Figure 2. 3. p201GCas9 targeting of *GmSHMT08* in soybean hairy roots. A. Each row represents *GmSHMT08* sequences in an independent hairy root modified by SHMT8-1. Wildtype (Forrest) is shown in the first row (green). Blue highlight represents the PAM site. Modified gRNA target sites are highlighted in yellow. **B.** The upper chromatogram is the sequencing result of *GmSHMT08* in an EXF67 hairy root transformed with the empty vector. The gRNA target site is outlined by a red box and PAM is highlighted by the blue box. The lower chromatogram is an example of modified *GmSHMT08* in an EXF67 hairy root transformed with SHMT8-1. All samples were sequenced using Forrest Sac1 or Tilling F primers.

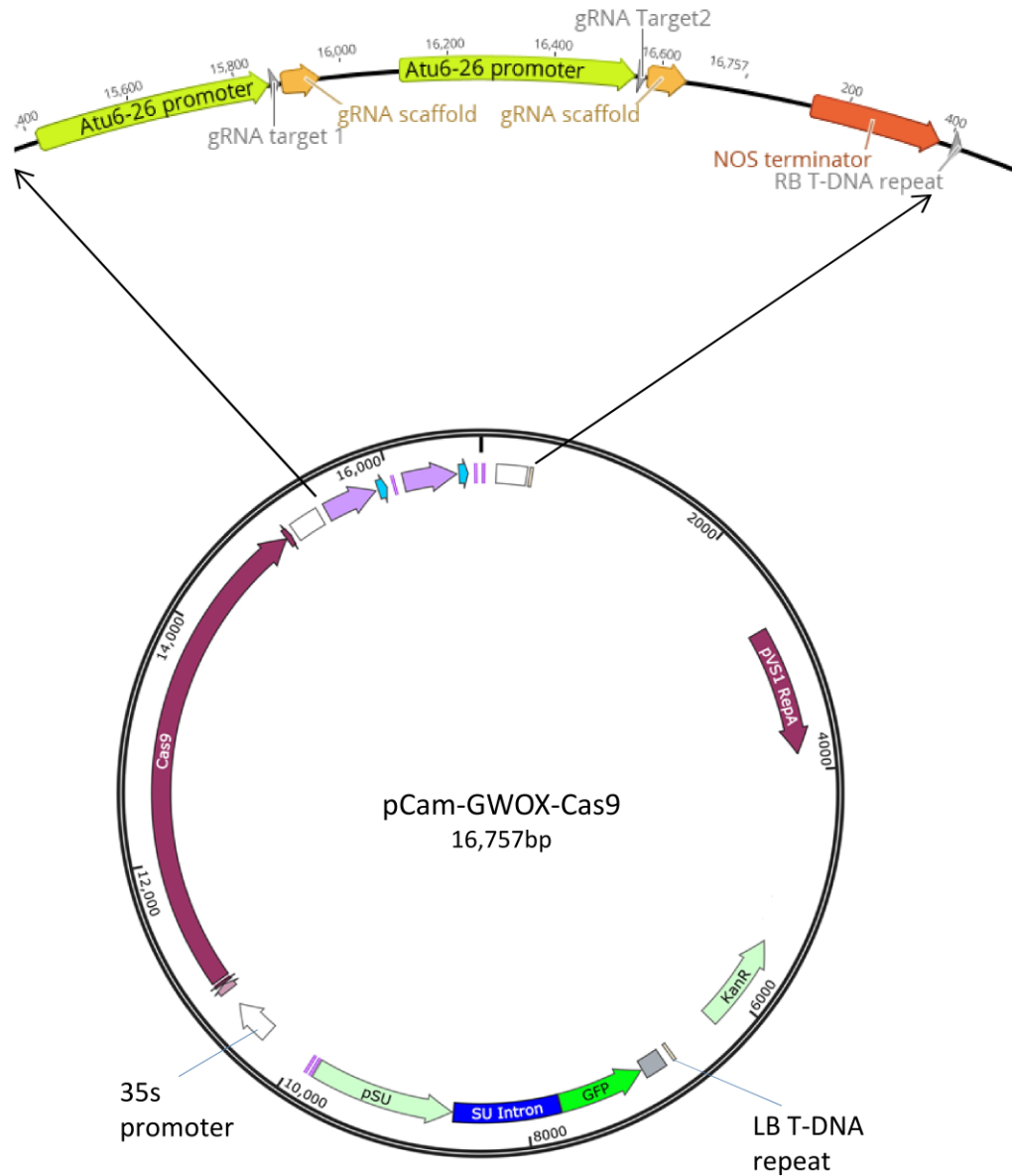


Figure 2. 4. Vector used to induce *GmSHMT08* deletions in hairy roots. The backbone pCamGFP-GWOX contained GFP for selection of transformed hairy roots and Kanamycin resistance for propagation in bacteria. *Streptococcus pyogenes* Cas9 (hspCas9) driven by CaMV 35s promoter is from the 35s-Cas9-SK vector. Two At6.26 promoters originally from Arabidopsis drive two different gRNAs which are inserted from the At6-26-SK vector.

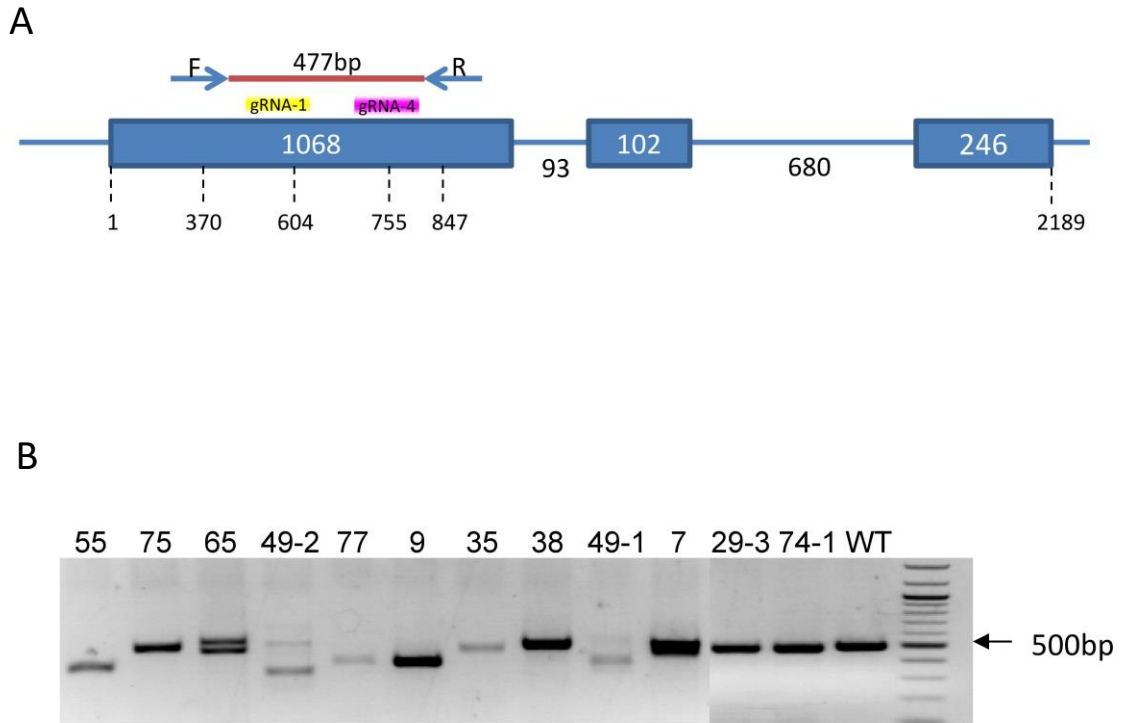


Figure 2. 5. Gel images of nucleotide deletions in *GmSHMT08*. **A.** Schematic model of *GmSHMT08*. Flanking primers (arrows labeled as F and R) were used to amplify the *GmSHMT08* coding region from the upstream of gRNA-1 to downstream of gRNA-4. **B.** PCR assay to detect SHMT8-14 CRISPR/Cas9-induced mutation in the target gene. Each lane represents PCR products of *GmSHMT08* amplified from gDNA extracted from independent hairy roots. WT = wild-type control hairy root transformed with BbSI vector.

A

```

>cDNA of GmSHMT08 (1416bp)
  1 atggatccag taagcgtgtg gggtaacacg cccttggcga cggtggtacc cgagatccat
 61 gacctcatcg agaaggagaa gcgcggtcaa tgccgcggaa tcgagctcat cgcctccgag
121 aacttcaact ccttcgcctg catcgaggcc ctcggcagcg ctctcacgaa caaatactcc
181 gagggcatgc cgggcaaccg ctactacggc ggcaatgaat acatcgacca gatcgaaaac
241 ctctgccgct cacgcgcctt ccaagccttc cacctcgacg cccaatcctg gggcgtcaac
301 gtccagccct actccggctc cccggccaac ttcgccgctt acaccgcctg cctcaacccc
361 cacgaccgca tcatggggct agatctccgc → tccggcggcc acctcaccca cggctactac
421 acctccggcg gaaagaagat ctccgccacc tccatttact tcgagagtct cccttacaag
481 gtaaaactca ccaccggcta catcgactac gaccgcttgg aagaaaaagc cctagacttc
541 aggcctaaac tcataatctg cggtggcagc gcgtaccctc gcgattggga ctacaaacgt
601 ttcaggggaag tcgctgataa gtgcggagca ttgcttctct gcgacatggc gcacactagc
661 ggccttgtgg ccgcgcagga agtgaacagc cccttcgagt attgcgacat tgtgaccacc
721 acgactcaca agagctttgcg ggggcccacgt gcgggggatga tcttttaccg gaagggcccc
781 aagccgccga agaaggggca gccggagaac gcggtttatg atttcgagga caagattaac
841 ttccggtgtg tcccttcgct gcagggtggg cccacaacc accagatcgg tgctctcgcc
901 gtggcgctga agcaggccgc gtcgccggg ttttaaggcct acgcgaagca ggtaaggcg
961 aacgcogttg cgcttgaaa atacttgatg gggaaagggt acagccttgt cactggcgga
1021 acggagaacc atcttgtttt gtgggatctg agaoctcttg gattgactgg gtataaggtg
1081 gagaaactct gtgatctctg taacattact gttacaaga acgctgtttt tggtgatagc
1141 agtgcccttg ccctgggtg agtgcgaaatt ggtgccctg ccataacttc taggggtttg
1201 gttgaaaaag actttgagca gattggtgag ttccttcacc gtgctgtgac tctcacactg
1261 gagatccaga aggagcatgg caaacttctc aaggatttca acaagggtct cgtaacaac
1321 aaggctattg aagatctcaa agctgatggt gagaagttct ctgccttgtt tgacatgcct
1381 ggcttctctg tatctgaaat gaagtacaag gatttag
  
```

B

Hairy root numbers	Examples of alignment with Forrest <i>GmSHMT08</i>
1	
2	
3	
4	
5	
6	
7	

C

```
>GmSHMT08 sequencing from a CRISPR/Cas9 hairy root line
NNNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNCAGCGCTCTCACGAACAAATNNCNCNCCGNANGGGCATG
CCGGGCAACCGCTACTACGGCGGCAATGAATACATCGACCAGATCGAAAACCTCNGCCGCTCACGCG
CCCTCCAAGCCTTCCACCTCGACGCCCAATCCTGGGGCGTCAACGTCCAGCCCTACTCCGGCTCCCC
GGCCAACTTCGCCGCCTACACCGCCGTCTCAACCCCCACGACCGCATCATGGGGCTAGATCTCCGC
TCCGGCGGCCACCTCACCCACGGCTACTACACCTCCGGCGGAAAGAAGATCTCCGCCACCTCCATTT
ACTTCGAGAGTCTCCCTTACAAGGTAAACTCCACCACGGCTACATCGACTACGACCGCTTGGAAGA
AAAAGCCCTAGACTTCAGGCCAAAACCTATAATCTGCGGTGGCAGCGCGTACCCTCGCGATTGGGAC
TACAAACGTTTCAGGGAAGTCGCTGATA  
NNNNNGNGNNNNNNNTTNNCNGNANGGNCCCNANNCNNN
NNANAANGGGNNCCNGNNNANNNNGNTTNNNAATNNNANGAANAAAATNANNTNNNNNNNNNNCNTN
NNNNNNNNNNNNNNNNNNNNCTTTTTNNNNNNNNNNNNCCNTTNNNNNNNNNNNNNNNNNNNNCGN
NNNNNNNNNNNCNNCCNNNNNNNNNNNNNNNNNNNTNNNNNNNNNNCANGNNNNNNNNNNNA
```

Figure 2. 6. Nucleotide deletion illustration and sequencing results. **A.** The position of gRNA-1 (yellow) and gRNA-4 (purple) on cDNA of Forrest type *GmSHMT08*. Green highlight represents nucleotides between two gRNAs. Forward and reverse primers used for DNA amplification are shown as black arrows. **B.** Sequencing results of multiple hairy roots transformed with SHMT8-14 and alignment with Forrest type *GmSHMT08* showing deletion in nucleotide sequence. **C.** An example of a sequencing result obtained for one hairy root line.

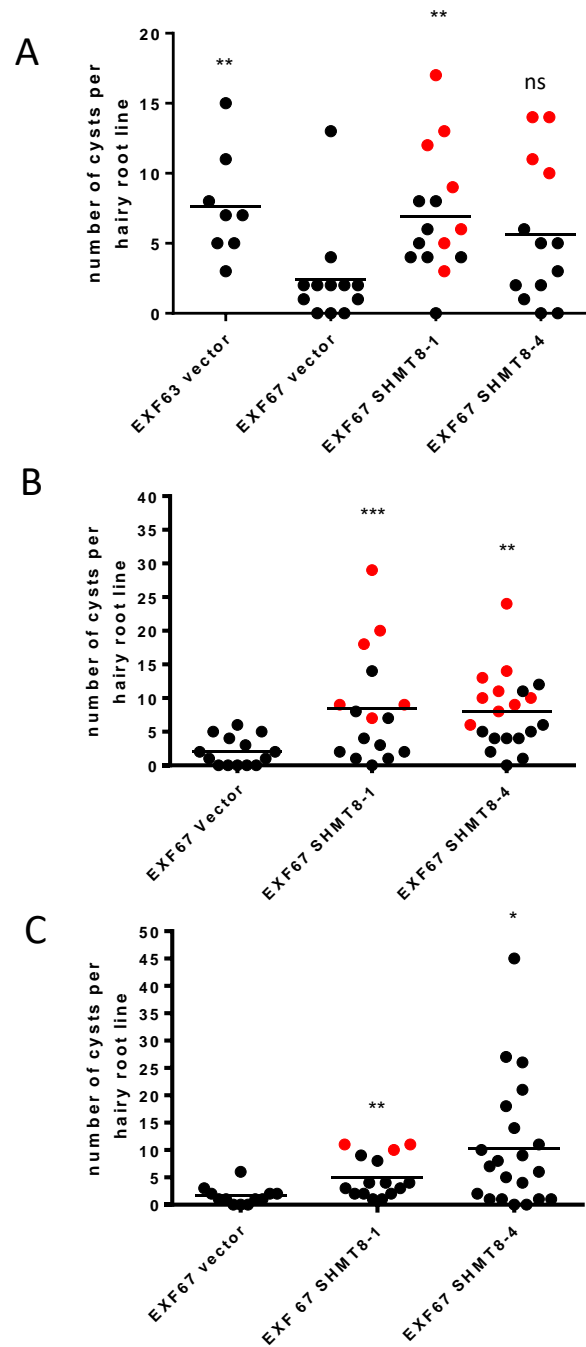


Figure 2. 7. Infection assay of EXF67 SHMT8-1 and SHMT8-4. A-C are three independent biological replicates of nematode development on EXF67 hairy root lines transformed with SHMT8-1 and SHMT8-4. Black dots represent the number of cysts on a single root system or hairy root line at 30dpi and the bars indicate the mean values. Red dots represent modified roots determined by sequencing. Asterisks indicate statistically significant differences compared to vector control as determined by a Student's *t*-test, * means $P < 0.05$, ** means $P < 0.01$, *** means $P < 0.001$, ns means not significant ($P > 0.05$).

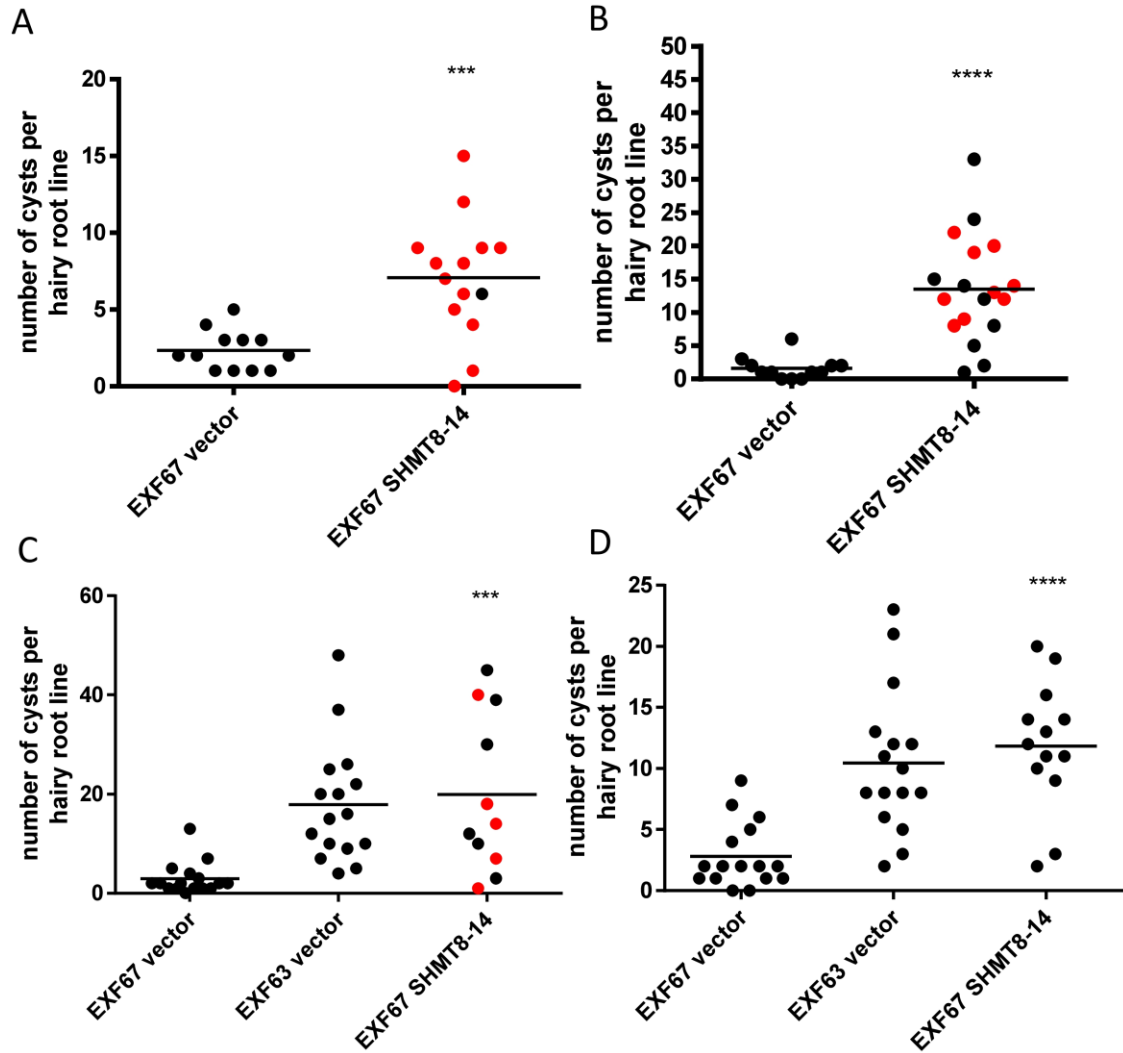


Figure 2. 8. Infection assay of EXF67 SHMT8-14. A-D are four independent biological replicates of nematode development on EXF67 hairy root lines transformed with SHMT8-14. Red dots in A represent confirmed nucleotide deletion on *GmSHMT08* by sequencing. Red dots in B and C were confirmed for nucleotide deletions on *GmSHMT08* with PCR/Gel image only. Note: Lines in D were not evaluated for modification. Asterisks indicate statistically significant differences compared to the ExF67 vector control as determined by a Student's *t*-test, *** means $P < 0.001$, **** means $P < 0.0001$).

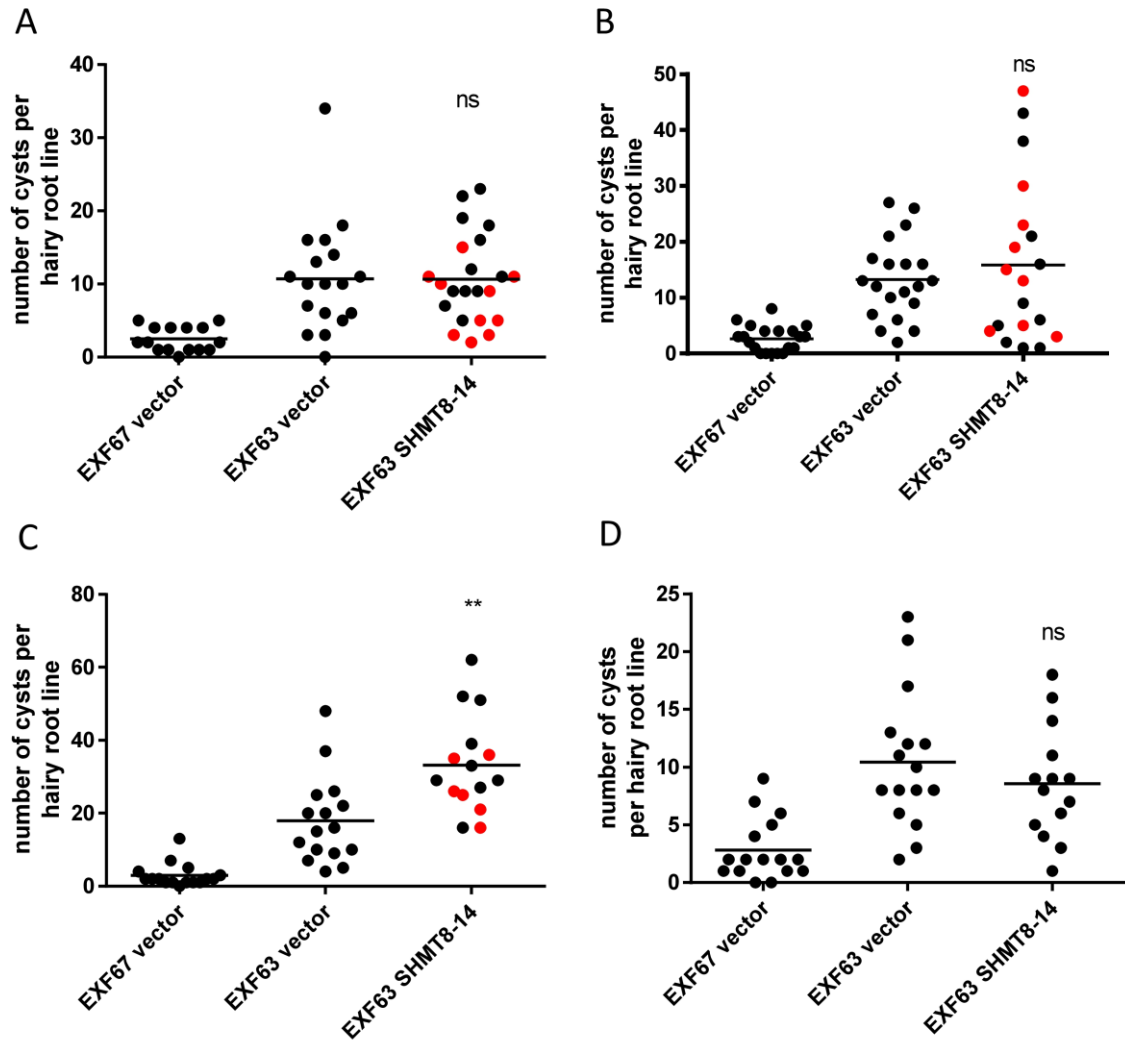


Figure 2. 9. Infection assay of EXF63 SHMT8-14. A-D are four independent biological replicates of nematode development on EXF63 hairy root lines transformed with SHMT8-14. Red dots in A, B and C represent confirmed nucleotide deletions in *GmSHMT08* by PCR/Gel image only. Roots in D were not evaluated for modification. Asterisks indicate statistically significant differences compared to EXF63 vector control as determined by a Student's *t*-test, ** means $P < 0.01$, ns means not significant ($P > 0.05$).

```

SHMT5      MDPVSVWGNTPLATVDPEIHDLEKEKH RQCRGIELIASENFTSFAVIEALGSALTNKYS
E-SHMT8    MDPVSVWGNTPLATVDPEIHDLEKEKR RQCRGIELIASENFTSFAVIEALGSALTNKYS
F-SHMT8    MDPVSVWGNTPLATVDPEIHDLEKEKR RQCRGIELIASENFTSFAVIEALGSALTNKYS
*****:*****

SHMT5      EGMPGNRYGGNEF IDQIENLCRSRALQAFHLDAQSWGNNVQPYSGSPANFAAYTAVLNP
E-SHMT8    EGMPGNRYGGNEY IDQIENLCRSRALQAFHLDAQSWGNNVQPYSGSPANFAAYTAVLNP
F-SHMT8    EGMPGNRYGGNEY IDQIENLCRSRALQAFHLDAQSWGNNVQPYSGSPANFAAYTAVLNP
*****:*****

SHMT5      HDRIMGLDL P SGGHLTHGYTSGGKKISATSIYFESLPYKVNSTTGYIDYDRLEEKALDF
E-SHMT8    HDRIMGLDL P SGGHLTHGYTSGGKKISATSIYFESLPYKVNSTTGYIDYDRLEEKALDF
F-SHMT8    HDRIMGLDL P SGGHLTHGYTSGGKKISATSIYFESLPYKVNSTTGYIDYDRLEEKALDF
***** *****

SHMT5      RPKLIICGGSAYPRDWDYKRFRE TADKCGALLLCDMAHTSGLVAAQEVNSPFEYCDIVTT
E-SHMT8    RPKLIICGGSAYPRDWDYKRFRE VADKCGALLLCDMAHTSGLVAAQEVNSPFEYCDIVTT
F-SHMT8    RPKLIICGGSAYPRDWDYKRFRE VADKCGALLLCDMAHTSGLVAAQEVNSPFEYCDIVTT
*****:*****

SHMT5      TTHKSLRGPRAGMI FYRKGPKPPKKGQPENAVYDFEDKINFAVFPSLQGGPHNHQIGALA
E-SHMT8    TTHKSLRGPRAGMI FYRKGPKPPKKGQPENAVYDFEDKINFAVFPSLQGGPHNHQIGALA
F-SHMT8    TTHKSLRGPRAGMI FYRKGPKPPKKGQPENAVYDFEDKINFAVFPSLQGGPHNHQIGALA
***** *****

SHMT5      VALKQAASPGFKAYAKQVKANAVALGN YLMGKGYSLVTGGTENHLVLWDLRPLGLTG NKV
E-SHMT8    VALKQAASPGFKAYAKQVKANAVALGK YLMGKGYSLVTGGTENHLVLWDLRPLGLTG NKV
F-SHMT8    VALKQAASPGFKAYAKQVKANAVALGK YLMGKGYSLVTGGTENHLVLWDLRPLGLTG YKV
*****:*****

SHMT5      EKLCDLCNITVNKNAVFGDSSALAPGGVRI GAPAMTSRGLVEKDFEQIGEF LHRAVTLTL
E-SHMT8    EKLCDLCNITVNKNAVFGDSSALAPGGVRI GAPAMTSRGLVEKDFEQIGEF LHRAVTLTL
F-SHMT8    EKLCDLCNITVNKNAVFGDSSALAPGGVRI GAPAMTSRGLVEKDFEQIGEF LHRAVTLTL
***** *****

SHMT5      EIQKEHGKLLKDFNKGLVNNKAI EDLKADVEKFSAT FDMPGFLVSEMKYKD*
E-SHMT8    EIQKEHGKLLKDFNKGLVNNKAI EDLKADVEKFSAL FDMPGFLVSEMKYKD*
F-SHMT8    EIQKEHGKLLKDFNKGLVNNKAI EDLKADVEKFSAL FDMPGFLVSEMKYKD*
***** *****

```

Figure 2. 10. Amino acid alignment of GmSHMT05, Essex type GmSHMT08, and Forrest type GmSHMT08. GmSHMT05 contains a total of 471 amino acids. GmSHMT05 and Essex Type GmSHMT08 have five amino acid polymorphisms (yellow-blue). GmSHMT05 and Forrest type GmSHMT08 have seven amino acid polymorphisms (yellow-blue and yellow-purple). Essex type GmSHMT08 and Forrest type GmSHMT08 have two amino acid polymorphisms (yellow-purple).

Acknowledgements

This was supported by Missouri Soybean Merchandising Council and United Soybean Board funding to Melissa G. Mitchum. We also thank the Daniel F. Millikan Endowment for partial support for Jingwen Kang. We wish to thank Bob Heinz and Amanda Howland for maintenance of nematode cultures, Pramod Kandoth for sharing expertise and technical assistance, and Tom Jacobs, Wayne Parrott, Jian-Kang Zhu, Cuong X. Nguyen, and Gary Stacey for providing CRISPR vectors and helpful discussion.

Chapter 3: Summary and Future Perspectives

Currently, SCN is the number one pathogen of soybean and causes more than \$1.2 billion in yield loss annually in the United States (Koenning & Wrather, 2010). However, the management strategy of SCN is limited. There are currently no genetically modified soybean cultivars that are specifically developed for SCN resistance, while to avoid SCN infestation, most soybean growers choose natural SCN resistant soybean cultivars that are derived from USDA germplasm in traditional breeding programs. On the molecular level, the mechanism for how SCN resistance functions in soybean is still not fully understood. It is necessary to continue the study of molecular mechanisms underlying SCN resistance in soybean; therefore, biotechnology could be utilized to allow the development of novel strategies to enhance SCN resistance.

We previously used a combination of reverse genetic methods including TILLING, VIGS, and RNAi to demonstrate that a SHMT gene (*GmSHMT08*) at the *Rhg4* locus is responsible for SCN resistance in the Peking-type soybean cv. Forrest (Liu et al., 2012). The *GmSHMT08* gene, which encodes a ubiquitous enzyme called serine hydroxymethyltransferase (SHMT) present in all organisms, was identified as playing a role in Peking-type resistance at the *Rhg4* locus (Liu et al., 2012). However, due to the highly duplicated genome of soybean, the ability to identify mutants of interest can be masked by functional redundancy in the genome. In recent years, several studies have reported the successful use of the precise genome-editing tool CRISPR/Cas9 in soybean (Du et al., 2016, Michno et al., 2015, Cai et al., 2015, Sun et al., 2015, Jacobs et al., 2015). Therefore, we targeted the *GmSHMT08* gene, which has been demonstrated to play a role in SCN resistance to test the utility of two different CRISPR/Cas9 systems

and their application to our well-established soybean hairy root infection assay pipeline.

Two constructs (SHMT8-1 and SHMT8-4) utilizing two different single gRNAs have proven to be effective in editing *GmSHMT08* in EXF67 (RIL harboring Forrest type SCN resistance) soybean hairy roots, but several limitations have appeared in these experiments. For example, the modification rate was calculated based on confirmed modified sequences, while the chimeric tissues may affect the accuracy of modification rate. Hairy roots generated from the same cotyledons may harbor different types of insertion or deletions. Therefore, the Sanger sequencing results were not completely reliable as reference to pre-select hairy roots for further SCN J2 infection. To correctly draw the correlation between nematode reproduction and gene knockout, it requires further screening to validate the genotypes; still, some mixed mutations might be masked. Nevertheless, we found out that hairy root lines generated using either SHMT8-1 or SHMT8-4 showed a statistically significant increase in nematode infection compared to empty vector control roots in three biological replicates of the experiment, indicating the actual modification rate might be higher than what we calculated through sequencing. Additionally, off-targeting analysis showed no mutation on the closely related copy of *GmSHMT08* which indicated that the using CRISPR/Cas9 system could enhance specificity on genome editing.

On the other hand, the dual gRNA CRISPR/Cas9 system significantly increased modification efficiency through inducing various sizes of nucleotide deletions on the target gene. However, detecting mutations simply through PCR/gel images might underestimate the mutation efficiency. Through amplification of *GmSHMT08* and comparison of the shifts on gel images, EXF67 hairy roots transformed with SHMT8-14

showed around 60% nucleotide deletion on *GmSHMT08* which is relatively higher than modification efficiency of the single gRNA CRISPR/Cas9 constructs. It is worth mentioning that, the significance values and the average nematode infection level are much higher compared to the previous SHMT8-1 and SHMT8-4 experiment. Taken together, these data indicate that the transformation with SHMT8-14 could have made up to 90% of deletions and effective gene function disruptions in hairy roots. Since most hairy roots have different levels of nucleotide deletion, pre-screening with PCR/gel could be optional in further application of the SHMT8-14 construct.

We further utilized the SHMT8-14 construct to understand the function of *GmSHMT08* in SCN susceptible soybean, however, imperfect random or uncontrollable SCN infection in susceptible soybean could complicate the factors that altering the shift of susceptibility. Therefore, further investigations or approaches might be necessary to understand the underlying mechanism of *GmSHMT08* in SCN susceptible background. The SHMT8-14 hairy root infection assays suffer from poor infection rates leading to a high degree of root to root variability in cyst counts within a single genotype (e.g. EXF67 and EXF63). Interestingly, in the infection assay where a statistically significant increase in susceptibility was observed, the average number of cysts in both susceptible vector control and SHMT8-14 lines was 50% higher than the other three replicates which is indicative of a good infection rate. Taken together, the infection conditions need to be optimal for maximum numbers of SCN to complete their life cycle. Otherwise, we cannot conclude that disruption of the Essex-type *GmSHMT08* increases susceptibility of soybean to SCN merely through hairy root pipeline.

Two single gRNA CRISPR/Cas9 constructs (SHMT5 and SHMT58) were applied

in both EXF63 and EXF67 backgrounds, but the two constructs may have the same drawbacks as that of SHMT8-1 or SHMT8-4. No obvious hairy root growth phenotypes were noted but hairy roots generated from EXF67 soybeans had severe bacteria contamination in the first two replicates. Transformation of EXF67 and EXF63 using SHMT5 did not alter the resistant phenotype in two biological replicates. The SHMT58 construct was also used to generate hairy roots in EXF67 soybeans, but unfortunately no *GmSHMT05* or *GmSHMT08* modifications were detected. Consequently, these replicates did not reveal a significant relationship of these two genes in plant growth and development.

Overall, the two types of CRISPR/Cas9 systems confirm our prior VIGS, RNAi, and TILLING results indicating that disruption of the Forrest type *GmSHMT08* gene function increases SCN susceptibility in SCN resistant soybeans. Moreover, we validate CRISPR/Cas9 as an efficient method to test soybean genes for a role in resistance to SCN using our transgenic hairy root pipeline. The dual gRNA CRISPR/Cas9 approach could achieve a higher percentage of mutation/deletion on the target gene.

Further research will be focused on designing dual gRNA CRISPR/Cas9 constructs targeting *GmSHMT05* and other candidate genes functioning in SCN resistance. Although the dual gRNA CRISPR/Cas9 system could significantly increase mutation/deletion efficiency, selecting efficient-editing gRNAs is still a critical process when designing CRISPR/Cas9 constructs. More details and insights on gRNA selection could benefit researchers to select proper gRNAs to achieve a high percentage of gene function disruption. Once gRNAs are highly specific and effective to knock out the target gene, the pre-screening of modified plant samples could be optional which could

potentially shorten the overall duration of the experiment. Other limitations that still need to be resolved include the methods used to screen plant tissue genotypes. Currently, either PCR/gel image or sequencing do not provide accurate modification rate, therefore increasing the gRNA efficiency is a priority. In the future, the CRISPR/Cas9 system may not be limited in application to the soybean hairy root pipeline. Additional methods of testing such as using the soybean composite plant system (Guo et al., 2011) and/or stable whole plant soybean transformation will be required to further evaluate a role for *GmSHMT08* in basal resistance.

Figure A.2. cDNA alignment of Essex type *GmSHMT08* (Sbjct) with *GmSHMT05* (Query).

	Score	Expect	Identities	Gaps	Strand
	2294 bits(1242)	0.0	1358/1416(96%)	0/1416(0%)	Plus/Plus
Query	1	ATGGATCCAGTGAGCGTGTGGGGTAACACGCCCTTGGCGACGGTGGATCCCGAGATCCAT	60		
Sbjct	1	ATGGATCCAGTAAGCGTGTGGGGTAACACGCCCTTGGCGACGGTGGATCCCGAGATCCAT	60		
Query	61	GACCTCATCGAGAAGGAGAAGCACCCGAATGCCGCGGCATCGAGCTCATCGCCTCCGAG	120		
Sbjct	61	GACCTCATCGAGAAGGAGAAGGCCCGTCAATGCCGCGGAATCGAGCTCATCGCCTCCGAG	120		
Query	121	AACTTCACCTCCTTCGCAGTCATCGAGGCCCTCGGCAGCGCCCTACCAACAAATACTCC	180		
Sbjct	121	AACTTCACCTCCTTCGCCGTCATCGAGGCCCTCGGCAGCGCTCTCACGAACAAATACTCC	180		
Query	181	GAGGGCATGCCCGCAACCGCTACTACGGCGGCAACGAATTCATCGACCAGATCGAAAAC	240		
Sbjct	181	GAGGGCATGCCCGCAACCGCTACTACGGCGGCAATGAATACATCGACCAGATCGAAAAC	240		
Query	241	CTCTGCCGCTCACGCGCCCTCCAGGCCCTTCCACCTCGACGCCCAATCCTGGGGCGTCAAT	300		
Sbjct	241	CTCTGCCGCTCACGCGCCCTCCAAGCCTTCCACCTCGACGCCCAATCCTGGGGCGTCAAC	300		
Query	301	GTGCAGCCCTACTCCGGCTCCCCTGCCAACTTCGCCGCCTACACCGCCGTCCTCAACCCC	360		
Sbjct	301	GTCCAGCCCTACTCCGGCTCCCAGGCCAACTTCGCCGCCTACACCGCCGTCCTCAACCCC	360		
Query	361	CACGACCGCATCATGGGGCTGGATCTCCCCCTCCGGCGGCCACCTCACCCACGGCTACTAC	420		
Sbjct	361	CACGACCGCATCATGGGGCTAGATCTCCCCCTCCGGCGGCCACCTCACCCACGGCTACTAC	420		
Query	421	ACCTCCGGCGGCAAGAAGATCTCCGCCACCTCCATTTACTTCGAGAGCCTCCCTTACAAA	480		
Sbjct	421	ACCTCCGGCGGAAAGAAGATCTCCGCCACCTCCATTTACTTCGAGAGTCTCCCTTACAAG	480		
Query	481	GTAAACTCCACCACCGGCTACATCGACTACGACCGCTTGGAGGAAAAGCCCTAGACTTC	540		
Sbjct	481	GTAAACTCCACCACCGGCTACATCGACTACGACCGCTTGGAGGAAAAGCCCTAGACTTC	540		
Query	541	AGGCCCAAATTGATAATCTGCGGTGGCAGCGCTTACCCTCGCGATTGGGACTACAAACGC	600		
Sbjct	541	AGGCCAAAATCATAATCTGCGGTGGCAGCGCTACCCTCGCGATTGGGACTACAAACGT	600		
Query	601	TTCAGGGAAATTGCGGATAAGTGCGGGGCATTGCTTCTCTGCGACATGGCGCACACTAGC	660		
Sbjct	601	TTCAGGGAAGTCGCTGATAAGTGCGGAGCATTGCTTCTCTGCGACATGGCGCACACTAGC	660		
Query	661	GGCCTTGTGGCCGCGCAGGAAGTGAACAGCCCTTCGAGTATTGCGACATTGTGACCACC	720		
Sbjct	661	GGCCTTGTGGCCGCGCAGGAAGTGAACAGCCCTTCGAGTATTGCGACATTGTGACCACC	720		
Query	721	ACCACGCATAAGAGCTTGCAGGGTCCACGGCGGGGATGATCTTTTACCGBAAGGGCCCC	780		
Sbjct	721	ACGACTCACAAGAGCTTGCAGGGGCCACGTGCGGGGATGATCTTTTACCGBAAGGGCCCC	780		
Query	781	AAGCCGCCGAAGAAGGGGCGAGCCGGAGAACCGGTTTATGATTTTCGAGGACAAGATCAAC	840		

References

- Acedo, J.R., Dropkin, V.H. and Luedders, V.D.** (1984) Nematode population attrition and histopathology of *Heterodera glycines*-soybean associations. *J. Nematol.* **16**, 48-56.
- Acharya, B., Lee, S., Mian, M.A.R., Jun, T.H., McHale, L.K., Michel, A.P. and Dorrance, A.E.** (2015) Identification and mapping of quantitative trait loci (QTL) conferring resistance to *Fusarium graminearum* from soybean PI 567301B. *Theor. Appl. Genet.* **128**, 827-838.
- Bekal, S., Craig, J.P., Hudson, M.E., Niblack, T.L., Domier, L.L. and Lambert, K.N.** (2008) Genomic DNA sequence comparison between two inbred soybean cyst nematode biotypes facilitated by massively parallel 454 micro-bead sequencing. *Mol. Genet. Genomics.* **279**, 535-543.
- Bekal, S., Niblack, T.L. and Lambert, K.N.** (2003) A chorismate mutase from the soybean cyst nematode *Heterodera glycines* shows polymorphisms that correlate with virulence. *Mol. Plant Microbe Interact.* **16**, 439-446.
- Belhaj, K., Chaparro-Garcia, A., Kamoun, S. and Nekrasov, V.** (2013) Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. *Plant Methods*, **9**. doi: 10.1186/1746-4811-9-39.
- Belhaj, K., Chaparro-Garcia, A., Kamoun, S., Patron, N.J. and Nekrasov, V.** (2015) Editing plant genomes with CRISPR/Cas9. *Curr. Opin. Biotechnol.* **32**, 76-84.
- Bernard RL, Noel GR, Anand SC, Shannon JG.** (1988) Registration of Fayette soybean. *Crop Sci.* **28**, 1028–1029.
- Bortesi, L. and Fischer, R.** (2015) The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnol. Adv.* **33**, 41-52.
- Cai, Y., Chen, L., Liu, X., Sun, S., Wu, C., Jiang, B., Han, T. and Hou, W.** (2015) CRISPR/Cas9-mediated genome editing in soybean hairy roots. *PLoS One.* **10**, e0136064.
- Concibido, V.C., Diers, B.W. and Arelli, P.R.** (2004) A decade of QTL mapping for cyst nematode resistance in soybean. *Crop. Sci.* **44**, 1121-1131.
- 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**, 1206-1209.

Dodds, P.N. and Rathjen, J.P. (2010) Plant immunity: towards an integrated view of plant–pathogen interactions. *Nat. Rev. Genet.* **11**, 539-548.

Endo, B.Y. (1986) Histology and ultrastructural modification induced by cyst nematodes. *Cyst Nematodes*. Springer, pp. 133-146.

Feng, Z., Mao, Y., Xu, N., Zhang, B., Wei, P., Yang, D.L., Wang, Z., Zhang, Z., Zheng, R., Yang, L., Zeng, L., Liu, X. and Zhu, J.K. (2014) Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. *Proc. Natl Acad. Sci.* **111**, 4632-4637.

Feng, Z., Zhang, B., Ding, W., Liu, X., Yang, D.L., Wei, P., Cao, F., Zhu, S., Zhang, F., Mao, Y. and Zhu, J.K. (2013) Efficient genome editing in plants using a CRISPR/Cas system. *Cell Res.* **23**, 1229-1232.

Gao, J., Wang, G., Ma, S., Xie, X., Wu, X., Zhang, X., Wu, Y., Zhao, P. and Xia, Q. (2015) CRISPR/Cas9-mediated targeted mutagenesis in *Nicotiana tabacum*. *Plant Mol. Biol.* **87**, 99-110.

Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A.H. and Moineau, S. (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, **468**, 67-71.

Gasiunas, G., Barrangou, R., Horvath, P. and Siksnys, V. (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *P. Natl Acad. Sci.* **109**, 2579-2586.

Gavassoni, W.L., Tylka, G.L. and Munkvold, G.P. (2001) Relationships between tillage and spatial patterns of *Heterodera glycines*. *Phytopathology*, **91**, 534-545.

Ghezzi, P., Garattini, S., Mennini, T., Bertini, R., Delgado Hernandez, R., Benigni, F., Sacco, S., Skorupska, M., Mengozzi, M., Latini, R., Kurosaki, M., Lombet, A., Fradin, A., Bonnet, J., Rolland, Y. and Brion, J.D. (1996) Mechanism of inhibition of tumor necrosis factor production by chlorpromazine and its derivatives in mice. *Eur. J. Pharmacol.* **317**, 369-376.

- Goverse, A. and Smant, G.** (2014) The activation and suppression of plant innate immunity by parasitic nematodes. *Annu. Rev. Phytopathol.* **52**, 243-265.
- Grant, D., Nelson, R.T., Cannon, S.B. and Shoemaker, R.C.** (2010) SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Res.* **38**, 843-846.
- Hartwig, E. F. & Epps, J. M.** (1973) Registration of 'Forrest' soybeans. *Crop Sci.* **13**, 287
- Holbein, J., Grundler, F.M. and Siddique, S.** (2016) Plant basal resistance to nematodes: an update. *J. Exp. Bot.* **67**, 2049-2061.
- Hsu, P.D., Ran, F.A., Scott, D.A., Lin, C.Y., Gootenberg, J.S., Konermann, S. and Zhang, F.** (2014) High precision genome editing by RNA-Guided CRISPR/Cas9. *Mol. Ther.* **22**, 216-216.
- Jacobs, T.B., LaFayette, P.R., Schmitz, R.J. and Parrott, W.A.** (2015) Targeted genome modifications in soybean with CRISPR/Cas9. *BMC Biotechnol.* **15**, 16.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E.** (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816-821.
- Jones, J.D. and Dangl, J.L.** (2006) The plant immune system. *Nature*, **444**, 323-329.
- Kim, J., Shirasawa, T. and Miyamoto, Y.** (2010) The effect of TAT conjugated platinum nanoparticles on lifespan in a nematode *Caenorhabditis elegans* model. *Biomaterials*, **31**, 5849-5854.
- 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.** (2010) Whole-genome sequencing and intensive analysis of the undomesticated soybean (*Glycine soja* Sieb. and Zucc.) genome. *Proc. Natl Acad. Sci.* **107**, 22032-22037.
- 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 Prog.* **10**, 1094.

Kunin, V., Sorek, R. and Hugenholtz, P. (2007) Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol.* **8**. doi: 10.1186/gb-2007-8-4-r61

Kyndt, T., Fernandez, D. and Gheysen, G. (2014) Plant-parasitic nematode infections in rice: molecular and cellular insights. *Annu. Rev. Phytopathol.* **52**, 135-153.

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

Lee, T.G., Kumar, I., Diers, B.W. and Hudson, M.E. (2015) Evolution and selection of *Rhg1*, a copy-number variant nematode-resistance locus. *Mol. Ecol.* **24**, 1774-1791.

Li, J.F., Norville, J.E., Aach, J., McCormack, M., Zhang, D., Bush, J., Church, G.M. and Sheen, J. (2013) Multiplex and homologous recombination-mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9. *Nat. Biotechnol.* **31**, 688-691.

Li, W., Teng, F., Li, T. and Zhou, Q. (2013) Simultaneous generation and germline transmission of multiple gene mutations in rat using CRISPR-Cas systems. *Nat Biotechnol.* **31**, 684-686.

Li, Y.H., Qi, X.T., Chang, R.Z. and Qiu, L.J. (2011) Evaluation and utilization of soybean germplasm for resistance to cyst nematode in China. *Soybean - Molecular Aspects of Breeding*. Croatia, Rijeka: InTech.

Liang, Z., Zhang, K., Chen, K. and Gao, C. (2014) Targeted mutagenesis in *Zea mays* using TALENs and the CRISPR/Cas system. *J. Genet. Genomics*, **41**, 63-68.

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

Liu, X., Liu, S., Jamai, A., Bendahmane, A., Lightfoot, D.A., Mitchum, M.G. and Meksem, K. (2011) Soybean cyst nematode resistance in soybean is independent of the *Rhg4* locus *LRR-RLK* gene. *Funct. Integr. Genomics*, **11**, 539-549.

- Mao, Y., Botella, J.R. and Zhu, J.K.** (2016) Heritability of targeted gene modifications induced by plant-optimized CRISPR systems. *Cell Mol. Life Sci.* doi:10.1107/s00018-016-2380-1.
- Mao, Y., Zhang, H., Xu, N., Zhang, B., Gou, F. and Zhu, J.K.** (2013) Application of the CRISPR-Cas system for efficient genome engineering in plants. *Mol. Plant*, **6**, 2008-2011.
- 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 *Rhg1* and *Rhg4* loci. *Theor. Appl. Genet.* **103**, 710-717.
- Mitchum, M.G.** (2016) Soybean resistance to the soybean cyst nematode *Heterodera glycines*: an update. *Phytopathology*, **12**, 1444-1450.
- Mitchum, M.G., Wang, X. and Davis, E.L.** (2008) Diverse and conserved roles of CLE peptides. *Curr. Opin. Plant Biol.* **11**, 75-81.
- Muthamilarasan, M. and Prasad, M.** (2013) Plant innate immunity: An updated insight into defense mechanism. *J. Bioscience*, **38**, 433-449.
- Peng, H.C. and Kaloshian, I.** (2014) The tomato leucine-rich repeat receptor-like kinases SISRK3A and SISRK3B have overlapping functions in bacterial and nematode innate immunity. *PLoS One*, **9**, e93302.
- Podevin, N., Davies, H.V., Hartung, F., Nogue, F. and Casacuberta, J.M.** (2013) Site-directed nucleases: a paradigm shift in predictable, knowledge-based plant breeding. *Trends Biotechnol.* **31**, 375-383.
- Riggs, R.D.** (1977) Worldwide distribution of soybean-cyst nematode and its economic importance. *J. Nematol.* **9**, 34-39.
- Sandermann, H., Jr., Schmitt, R., Eckey, H. and Bauknecht, T.** (1991) Plant biochemistry of xenobiotics: isolation and properties of soybean O- and N-glucosyl and O- and N-malonyltransferases for chlorinated phenols and anilines. *Arch. Biochem. Biophys.* **287**, 341-350.
- Schmitt, D.P. and Sipes, B.S.** (2004) Nematode management in crops grown in North America and Hawaii. *Nematol. Monogr. Persp.* **2**, 63-70.

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**, 178-183.

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., Zhang, K., Liu, J., Xi, J.J., Qiu, J.L. and Gao, C. (2013) Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat. Biotechnol.* **31**, 686-688.

Shen, B., Zhang, W.S., Zhang, J., Zhou, J.K., Wang, J.Y., Chen, L., Wang, L., Hodgkins, A., Iyer, V., Huang, X.X. and Skarnes, W.C. (2014) Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat. Methods*, **11**, 399-402.

Smith, T. J. & Camper, H. M. Registration of 'Essex' soybean. (1973) *Crop Sci.* **13**, 495.

Song, Q.J., Marek, L.F., Shoemaker, R.C., Lark, K.G., Concibido, V.C., Delannay, X., Specht, J.E. and Cregan, P.B. (2004) A new integrated genetic linkage map of the soybean. *Theor. Appl. Genet.* **109**, 122-128.

Sorek, R., Lawrence, C.M. and Wiedenheft, B. (2013) CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annu. Rev. Biochem.* **82**, 237-266.

Terns, M.P. and Terns, R.M. (2011) CRISPR-based adaptive immune systems. *Curr. Opin. Microbiol.* **14**, 321-327.

Tylka, G.L., Marett, C.C. (2000) Evaluation of soybean varieties resistant to soybean cyst nematode in Iowa: 1999. *Publication IPM*. University Extension, Iowa State University: **52**.

Upadhyay, S.K., Kumar, J., Alok, A. and Tuli, R. (2013) RNA-guided genome editing for target gene mutations in wheat. *G3*, **3**, 2233-2238.

Wang, T., Wei, J.J., Sabatini, D.M. and Lander, E.S. (2014) Genetic screens in human cells using the CRISPR-Cas9 system. *Science*, **343**, 80-84.

- Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C. and Qiu, J.L.** (2014) Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat. Biotechnol.* **32**, 947-951.
- Warnke, S.A., Chen, S.Y., Wyse, D.L., Johnson, G.A. and Porter, P.M.** (2008) Effect of rotation crops on hatch, viability and development of *Heterodera glycines*. *Nematology*, **10**, 869-882.
- Weigel, D. and Glazebrook, J.** (2002) *How to Transform Arabidopsis*, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Wrather, J., Anand, S. and Dropkin, V.** (1984) Soybean cyst nematode control. *Plant Dis*, **68**, 829-833.
- Wu, X.Y., Zhou, G.C., Chen, Y.X., Wu, P., Liu, L.W., Ma, F.F., Wu, M., Liu, C.C., Zeng, Y.J., Chu, A.E ang, Y.Y., Chen, J.Q. and Wang, B.** (2016) Soybean cyst nematode resistance emerged via artificial selection of duplicated serine hydroxymethyltransferase genes. *Front Plant Sci.* **7**, 998.
- Xie, K.B. and Yang, Y.N.** (2013) RNA-guided genome editing in plants using a CRISPR Cas System. *Mol. Plant*, **6**, 1975-1983.
- Zhang, H., Feng, R., Li, J., Guan, D. and Cao, F.** (2010) Genetic diversity and phylogeny of soybean rhizobia isolated from the regions of Loess Plateau in China. *Wei Sheng Wu Xue Bao*, **50**, 1466-1473.
- Zhou, J., Shen, B., Zhang, W., Wang, J., Yang, J., Chen, L., Zhang, N., Zhu, K., Xu, J., Hu, B., Leng, Q. and Huang, X.** (2014) One-step generation of different immunodeficient mice with multiple gene modifications by CRISPR/Cas9 mediated genome engineering. *Int. J. Biochem. Cell Biol.* **46**, 49-55.