Global DNA Methylation Changes in Soybean as a response to Soybean Cyst Nematode (SCN) infection.

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Aditi Rambani
May 2018

I dedicate this work to my dear friend Krishna and my parents. Their unconditional love and blessings made it all possible.

Acknowledgements

I feel immense gratitude towards everyone who helped me with this project. This project would not have been possible without the help and support of all the kind people around me, to only some of whom it is possible to give particular mention here. Foremost, is my advisor Dr. Tarek Hewezi, I am very thankful for his guidance, encouragement and patience. I sincerely appreciate the direction and input my committee members Dr. Staton, Dr. Pantalone, Dr. Tessa Burch Smith and Dr. Neal Stewart provided for this project. I would also like to thank my lab members Valeria Lopes, Dr. Yinfeng Zhu, Dr. Sarbottam Piya, Yanfeng (Ryan) Hu, Songnan Yang, John Hollis Rice, Jinyi Liu, Jordan Skeen, Megan Bruce, Kelsey Ritter, Morgan Bennett and Daniel Niyikiza. A special thanks to Thomas Lane and Priya Ranjan who extended help for bioinformatics analysis. I would also like to thank my fellow students and staff of Department of Plant Sciences at University of Tennessee for their cooperation and help throughout my project. Finally, I would like to thank my family for constant support that helped me maintain right perspective and attitude towards my work.

Abstract

Soybean cyst nematode (SCN, *Heterodera glycines*) is the most damaging pest in soybean production worldwide. The compatible interaction of SCN with susceptible soybean plants is mediated by differential expression of thousands of genes in the infected root cells, leading to the formation of a functional feeding site, the syncytium. During an incompatible interaction of SCN with resistant soybean, the developing syncytium degenerates leading to nematode death. The resistance to SCN in soybean is derived from two major loci, Rhg1 and Rhg4. In this study, the role of genome-wide DNA methylation in regulating gene expression during the compatible interaction was examined using susceptible soybean cultivar 'Williams 82'. The analysis revealed that SCN induces both hyper- and hypomethylation in thousands of genomic regions overlapping with genes. The level and pattern of DNA methylation in various genic regions were found to impact gene transcription. A significant number of the differentially methylated genes was found to overlap with genes known to be significantly differentially expressed in syncytium, providing the first experimental evidence that syncytium transcriptome is epigentically controlled. In addition, the levels and patterns of DNA methylation were compared between the compatible and incompatible interactions using a pair of near-isogenic lines differing in *Rhg4* allele. The methylomes of two near isogenic lines, susceptible (TN09-16) and the resistant (TN09-29) were substantially different both under SCN-infected and non-infected conditions. Stably heritable as well as novel non-parental differentially methylated regions in genes with functions related to SCN parasitism of soybean were discovered. Furthermore, differential DNA methylation in microRNA genes was also examined in the susceptible and resistant lines in response to SCN infection. A number of differentially methylated microRNAs were identified specifically in the susceptible lines, Williams 82 and TN09-16, indicating that various components of epigenetic mechanisms are mutually linked. Taken together, profiling DNA methylation at single nucleotide resolution during the susceptible and resistant interactions provided unprecedented insights into the role of this epigenetic mark in determining the compatibility of the interaction between

soybean and SCN by impacting the expression of protein-coding and microRNA genes as well as transposable elements located nearby genes.

Table of contents

List of figures

Chapter 1: General introduction and literature review

1. Introduction

Nematodes are unsegmented microscopic roundworms that inhabit a wide range of habitats. They contribute to many ecosystem functions, including nutrient cycling and biological control (Ferris, 2010). Free-living *Caenorhabditis elegans* has become increasingly popular as a model organism in the fields of neurobiology (Ankeny, 2001; Sengupta and Samuel, 2009) and animal-microbial symbioses (Sifri et al., 2005). Other pathogenic species affect billions of plants, animals and people across the globe. The constitutive presence of diverse nematode population in soils is particularly threatening to crops and livestock which make up our food supplies. Disease outbreak of the many agronomically-relevant nematode species not only cripples plant productivity, but devastates entire fields (Nicol et al., 2011). Plant parasitic nematodes (PPNs) may feed on roots, stem, foliage, and flowers using a specialized parasitism structure called the stylet. The stylet is the main morphological feature that distinguishes PPNs and allows for parasitism. This structure is a long and hollow, yet rigid spear-shaped protrusion in the anterior part of the nematode (mouth). Upon infection, PPNs use their stylet first to penetrate the plant cell wall and gain access to water and nutrients being transported throughout the plant. A small hole at the tip of the stylet provides an entry and exit point for the animal to exchange pathogenic effector proteins with host-derived nutrients. Effectors compromise host cell functions and help redirect nutrients, ultimately resulting in disease.

While there is a diverse array of plant–parasite dynamics, PPNs are broadly categorized in three groups based on their feeding habits. Endoparasites burrow inside host tissue for feeding, whereas ectoparasites feed from the surface of the tissue. Semi ecto-endo parasites may exhibit either behavior. PPNs are further categorized as sedentary or migratory, based on nematode mobility during feeding. Migratory ectoparasites move through soil and only temporarily stop to feed on plants in their path. Migratory endoparasites, on the other hand, enter the plant and feed as they move from cell to cell, causing immense cellular necrosis. Most migratory endoparasitic nematodes are obligate parasites and feed on root cortical cells or aerial parts of the plant. Ectoparasitic nematodes may also sedentarily feed on cortical and epidermal cells at the

same location for an extended period while remaining outside the plant (Tytgat et al., 2000). Semi ecto-endo nematodes feed like migratory ecto-parasites during certain life stages, but can also infiltrate the host completely for feeding. When sessile, semi ecto-endoparasites induce a feeding structure within the host plants. Permanent feeding structures are initiated by the sedentary endoparasitic nematodes to support their development and maturation (Davis et al., 2004). It is in these tumor-like feeding structures that plant resources are appropriated for nematode use. Plant cells are dramatically modified by effector proteins to facilitate feeding site formation and import nutrients for nematode growth and development (Hewezi and Baum, 2013). Through this interaction, sedentary endoparasitic cyst and root-knot nematodes are the most economically important species in agriculture. Nematode secretions are synthesized in the three esophageal gland cells; two subventral and one dorsal cell (Davis et al., 2008). Subventral glands are expected to be active throughout the life cycle of migratory nematodes (Jones et al., 2014) In sedentary endoparasites, however, subventral glands are active during the early stage of root penetration, migration and initiation of the feeding site (Hewezi and Baum, 2013). In contrast, the single dorsal gland cell is active during feeding site formation and nematode feeding (Gheysen and Mitchum, 2011).

2. Soybean Cyst Nematodes

Soybean cyst nematode (SCN) is a soil-dwelling obligate sedentary endoparasite. The life cycle of SCN progresses through four juvenile stages, each followed by a molt required to replace the outer cuticle. The first juvenile stage (J1) occurs inside the egg and can be sustained for long periods of time. In the presence of exogenous cues from nearby host plants, J1 molts to reach a second juvenile developmental stage (J2). This is the infective and motile stage of SCN, where J2s hatch and move toward the soybean root system. J2s penetrate root epidermis and migrate through cortical cells towards the vascular cylinder. Once reaching a suitable procambial cell near primary xylem, nematodes secrete effector proteins at the initial feeding site. The initial infected cell begins to rapidly change in cytoplasm density and cell wall structure. Protoplasts of the neighboring cells begin to fuse due to partial dissolution of cell walls. In this way, a

single feeding cell can merge with over 200 cells to form a large feeding structure, called the syncytium. The characteristic features of syncytium are enlarged nuclei, dense cytoplasm with heavy streaming, and high metabolic activity (Jones, 1981). Once the feeding site is initiated and established, SCN remains sessile as it matures into an adult. Gender of adult nematodes is determined by availability of food, where more males are generated in cases of low supply (Triantaphyllou, 1973). When food is abundant, both genders are produced and contribute to plant disease. Males briefly leave their feeding site approximately 15 days after infection to fertilize females and remain in soil thereafter (Triantaphyllou and Hirschmann, 1962). Females are large and bulbous compared to males and remain stationary throughout their life. In just 3-6 weeks after initial infection, females can lay hundreds of eggs inside its body causing her to form a bloated, lemon shape. After the mothers death, its body transforms into a tough cyst encapsulation around the eggs (Wyss and Zunke, 1986). Egg-containing cysts can remain viable in soil for about a decade under amicable environmental conditions (Inagaki and TsuTsuMI, 1971; Niblack et al., 2006).

Soybean (*Glycine max*) is the second most cultivated crop in the United States since its introduction in Georgia in 1765 (Hymowitz and Harlan, 1983). As a major source of edible and biodiesel oil, soybean demand has risen substantially in recent years and is expected to continue (Lamers et al., 2014). Soybean is also used in rotations between corn and wheat because of its interaction with nitrogen-fixing bacteria, *Bradyrhizobium japonicum,* to improve soil fertility. Heavy infestation of SCN can result in crop losses of up to 70% (Wrather and Koenning, 2006; Wrather and Koenning, 2009; Koenning and Wrather, 2010).. Below ground symptoms include fewer nitrogen-fixing nodules, more disease-related cysts, and root growth retardation. SCN pathogenesis also causes chlorosis and stunting in the above-ground portions of the plant. Together, these symptoms result in nutrition deficiencies that leave poorly productive plants. SCN populations continue to increase in density and genetic diversity proportional to increased soybean production. Vast abundance of genetic diversity in widespread populations and limited SCN resistant strategies pose a serious problem for one of the biggest crops of today (Wang et al., 2003).

2.1 SCN resistance in soybean

Several techniques have been developed to discover and understand the genetic basis of resistance in soybean. Soybean genomics research has produced many candidate genes for SCN resistance in the past decade. Soybean resistance to different SCN races (HG types) is known to be conferred by multiple quantitative trait loci (QTL) (Concibido et al., 2004). Yet, candidate gene discovery associated with large effect QTLs was stalled for several years because leucine rich transmembrane receptor like kinase (LRR-RLK) genes were thought to be the likely source of SCN resistance gene. In the case of tomato and potato cyst nematodes, resistance is mediated by coiled coil nucleotide binding-leucine-rich repeat (CC_NB_LRR) and toll-interleukin 1 receptors-nucleotide binding-leucine-rich repeat (TIR-NB-LRR) classes (Kandoth and Mitchum, 2013). However, these genes were discredited as candidate genes in soybean. Therefore, new methodologies were adopted to pursue this cause. Broadly, these methodologies can be divided into two groups, genetics approaches, and SCN-soybean interaction-based approaches.

2.1.1 Genetics Approaches

Classical genetic analysis has been utilized extensively to identify genetic control of SCN resistance in soybean (Rao-Arelli and Anand, 1988; Anand and Rao-Arelli, 1989). While several putative QTLs have been suggested, 'Rhg1' and 'Rhg4' have been repeatedly mapped in resistant genotypes. Rhg1 allele from PI 88788 designated as 'rhg1-b' is the prevalent source of SCN resistance in commercially available soybean in the United States. However, employment of this single strategy has led to the development of SCN tolerance (Adee et al., 2008). Fine mapping of rhg1-b allele and silencing of the LRR-RK gene at rhg1 locus using artificial microRNA provided evidence to depose LRR-RK as the candidate for SCN resistance at the Rhg1 locus (Kim et al., 2010; Melito et al., 2010). The Rhg1 associated region was later narrowed down to 31.2kb genomic segment, containing three complete genes (Cook et al., 2012). Copy number of the 31.2kb genomic

region varies among soybean accessions (Cook et al., 2012). Based on sequence homology, these three genes code for an amino acid transporter protein, α-SNAP or N-ethylmaleimide-sensitive factor attachment protein, and a wound inducible protein (WI12). 'Williams 82' is a susceptible soybean cultivar, which carries a single copy of the genes identified at Rhg1 locus, whereas resistant accessions contain up to 10 copies. As a trend, increased dosage of these three genes in soybean corresponds to higher resistance against SCN. This was further validated when a susceptible cultivar became partially resistant to SCN by overexpression of the three Rhg1 genes (Cook et al., 2012).

In addition to copy number variations, soybean cultivars display three distinct versions of Rhg1 gene sequences, Peking-type (P), PI88788-Fayette type (F), and Williams 82-type (W) (Cook et al., 2014; Lee et al., 2015). In summary, Rhg1 locus contains 3 genes within the 31.2kb region, which are inherited as a one unit with varied copy number. For example, the cultivar 'Fayette' carries a total of 10 tandem copies, nine of which are F type and one copy is W type. Along with copy number variation and sequence type, plant DNA methylation is a controlling factor of SCN resistance. Lines containing a single copy of Rhg1 locus were found to be hypomethylated when compared to resistant lines containing three copies. DNA methylation patterns of Rhg1 locus are stably inherited (Cook et al., 2014). Rhg1 locus-derived SCN resistance is further complexed by its interaction with another major SCN resistance locus, Rhg4 mapped to chromosome 8.

Epistatic interactions between Rhg1 and Rhg4 were first identified in 'Forrest', which was developed by introgression of Peking-derived SCN resistance into a high yielding variety (Meksem et al., 2001). Later, Lee et al. (2015) confirmed that cultivars carrying few copies of the Rhg1 locus can only confer SCN resistance in conjunction with the resistant allele at Rhg4 locus. High copy number of Rhg1 locus can confer resistance independent of the Rhg4 locus (Lee et al., 2015). These results confirmed that not LRR-RLK, but Rhg1 and 4 were the likely candidates for SCN resistance (Liu et al., 2011). Positional cloning of Rhg4 locus in three RILs resolved the interval to an 8kb region that contained only two genes. Of these, Serine hydroxyl methyltransferase (SHMT) was validated as the SCN resistance gene through targeted RNA interference (RNAi) and

virus-induced gene silencing (Liu et al., 2012). SHMT is highly conserved across many species of mammals and plants, but its function in nematode resistance remains completely unknown. Serine hydroxyl methyl transferase is an enzyme that catalyzes reversible conversions of serine and glycine. One-carbon units generated in this process are required for cellular metabolic processes such as purine biosynthesis, production of thymine nucleotides, and formation of S-adenosyl methionine (SAM). SHMTs are involved in several other cellular processes like DNA synthesis, which requires thymine nucleotides and methylation catalysis, which involve the activity of SAM as methyl donor.

2.1.2 SCN-soybean interaction-based approach

In resistant soybean, nematode secreted effector proteins are recognized by the host plant cells and the feeding sites are degenerated, leading to nematode developmental arrest. On the other hand, during compatible interactions, nematodes successfully induce a metabolically-active feeding site where the animals complete their life cycle and propagate (Klink et al., 2007; Klink et al., 2009; Klink et al., 2010). Over the last few years, many studies have documented changes in gene expression during various stages of syncytium development and nematode parasitism, providing key insights into molecular mechanisms that regulate soybean-SCN interactions. Before widespread use of RNA sequencing, most differential expression studies were conducted using a microarray platform. Custom soybean cDNA microarrays were used to assess changes in gene expression during SCN infection (Khan et al., 2004; Alkharouf et al., 2006). While it was clear that nematode infection induced higher expression of wound, stress and defense response genes in soybean during the compatible interaction, early microarray studies only analyzed a portion of the genome. More comprehensive gene expression quantification during the compatible and incompatible interactions was conducted using the Affymetrix Soybean Genome Array GeneChip , which contains probes for more than 35,000 soybean transcripts (Klink et al., 2007, 2007, 2010; Mazarei et al., 2011). Multiple reports agree that during the initial phases of SCN infection, soybean genes involved in defense, wound, and stress response are the most profoundly differentially

expressed. At later stages of infection, SCN modulates the expression of a significant number of genes involved in various cellular processes to maintain a functional feeding site. It is in the syncytium that many of the key molecular changes induced by SCN occur. To precisely follow these changes, experiments were designed to isolate syncytial cells using laser capture microdissection (Ithal and Mitchum, 2011). These studies portray transcriptome changes that take place during syncytium formation and provided a list of genes significantly differentially expressed between syncytium and root tissues.

Still, the genetic basis of SCN resistance cannot be resolved by simply observing gene expression changes during SCN-soybean interactions. Our picture of the molecular basis for soybean resistance is not complete without an understanding of the epigenetic mechanisms that accompany genetic regulation. A comprehensive analysis studying expression changes and underlying epigenetic factors that regulate gene expression can provide more extensive insight into the SCN-soybean interaction. Major breakthroughs have been made in deciphering which genetic locus confers resistance and how gene expression is different in resistant versus susceptible lines. The next step is to augment these findings with epigenetic regulation, whether by directly influencing expression at resistance locus like Rhg1 (Cook et al., 2014) or by changing global gene expression (Rambani et al., 2015).

3. Epigenetic Modifications

Epigenetic factors can be categorized into three groups– histone modifications, DNA methylation, and small RNA regulators.

3.1 Histone Modifications

Nucleosome are octamers consisting of two copies of H2A, H2B, H3, and H4 histone types each joined together by linker histones, H1 (Luger et al., 1997; Thomas, 1999;

Luger et al., 2012). Nucleosome architecture defines chromatin structure and accessibility to DNA transcriptional machinery. Histone type variation and monomer modifications both have monumental effects on gene transcription. Although nucleosome organization is canonical and ubiquitous across species, slight variations can be introduced by exchange of histone variants (Talbert and Henikoff, 2016). Histone variants differ slightly in amino acid sequence, but are effectively interchangeable in the octamer. For example, Arabidopsis histone variants H3.3, H2A, H2AZ, and H1.3 are present in transcriptionally active regions with high nucleosome turn over. In the same way, H3.1, H2AX, H2AW, H1.1, and H1.2 localize to heterochromatin and centromeric histone3 (CENH3) is almost exclusively incorporated within centromeres and pericentromeric regions (Zilberman et al., 2008; Talbert and Henikoff, 2010; Stroud et al., 2012; Wollmann et al., 2012; Shu et al., 2014; Talbert and Henikoff, 2014; Yelagandula et al., 2014; Steiner and Henikoff, 2015; McKinley and Cheeseman, 2016).

Modifications in chromatin structure near promoter regions changes its accessibility to transcription factors, resulting in differential gene expression in response to various developmental and environmental stimuli. Posttranslational modifications of histones introduce transient transcriptional changes by directly interacting with other epigenetic factors, such as small RNAs and DNA methylation, and by modulation how tightly DNA is wrapped around histones. Epigenetic modifications of histones can occur at different levels, providing a wide range of options for fine-tuning gene expression. Chemical modifications like methylation, acetylation, phosphorylation, and ubiquitination commonly occur at lysine and arginine residues of N-terminal histone tails to different degrees (mono-, di-, and tri-) (Bergmüller et al., 2007; Zhang et al., 2007; Roudier et al., 2011; Sequeira-Mendes et al., 2014; Mahrez et al., 2016). Associations between positive histone tails and negatively charged DNA is somewhat neutralized by negative charges on acetylated lysine, causing associated DNA to loosen and become more accessible (Struhl, 1998; Bannister and Kouzarides, 2011). Alternatively, histone tail modifications can also aid in recruiting elements that directly modify chromatin and DNA.

3.2 Small RNA Regulators

In plants, small RNAs are known to regulate gene expression during various cellular processes important for development, reproduction, and stress responses (Weiberg et al., 2014). This mechanism initially evolved to control transposition of active transposable elements (TEs) that destabilize genome organization. There is vast type and functional diversity small RNAs in plants. A simple, straightforward method has been proposed for small RNA categorization based on precursors (Axtell, 2013). Small RNAs derived from double-stranded RNA (dsRNA) are classified as small interfering RNAs (siRNAs), and those formed by processing single-stranded RNA hairpins are called hairpin RNAs (hpRNAs). siRNAs can be further divided into subclasses of heterochromatic siRNAs, secondary siRNAs, and natural antisense transcript siRNAs (NAT-siRNAs). The hpRNAs subclasses include microRNAs (miRNAs) that are transcribed by Pol II from endogenous miRNA genes and other hpRNAs that arise from, for example, Pol II transcription of palindromic sequences within active TEs (Slotkin et al., 2005). Both hpRNAs and siRNAs perform post transcriptional gene silencing (PTGS), but siRNAs can further establish transcriptional gene silencing (TGS) by mediating chromatin and epigenetic modifications on DNA (Rowley et al., 2017).

hpRNAs and siRNAs PTGS can effectively combat retrovirus infection. PGTS is initiated by the shredding of dsRNAs by DICER-LIKE 2 and 4 (DCL1 in case of hpRNAs) into small, 21-22nt pieces, called primary siRNAs. Primary siRNAs become single stranded and associate with Argonaute 1 (AGO1) to serve as a guide to target mRNAs by complementary base pairing. Targets of mRNAs are cut at the complementary sites by the endonuclease activity of AGO1, then RNA DEPENDENT RNA 6 (RDR6) polymerizes a complementary strand for the cleaved mRNA transcripts. Because dsRNA is the final product, PTGS may occur repeatedly by associating with DCL proteins (Figure 1). TGS is achieved by RNA-directed DNA methylation, where siRNAs guide methyltransferases to target loci for establishment of the repressive DNA methylation marks. Plant-specific RNA polymerase IV (PolIV) transcribes ssRNA from genomic regions harboring DNA

and histone epigenetic modifications. These ssRNAs are converted into dsRNA by RNA DEPENDENT RNA 2 (RDR2) polymerase and supplied to DCL3 to be diced into 24nt siRNAs (Onodera et al., 2005; Pikaard et al., 2008; Wierzbicki et al., 2008). Single strands of these 24nt siRNAs are picked up by AGO4 or 6 and recruited to the target loci. DNA methyltransferases are also recruited to the target loci, which catalyze methyl group attachment (Cao and Jacobsen, 2002; Zilberman et al., 2003; Wierzbicki et al., 2008; Wierzbicki et al., 2009; Zhang and Zhu, 2011; Borges and Martienssen, 2015; Cuerda-Gil and Slotkin, 2016).

3.3 DNA Methylation

The most well-known epigenetic modification is methylation of fifth carbon in cytosine (C) (Lister et al. 2009). In plant, cytosine methylation occurs in three sequence contexts—CG, CHG, and CHH, where H denotes any base other than guanine (G). DNA methylation is important for defining heterochromatic regions and centromeres as well as maintaining TEs dormancy (Law and Jacobsen, 2010). Methylation in promoter regions near the transcriptional start site is known to be negatively correlated with gene expression levels. Heavily methylated TEs also impact the expression of nearby genes (Hollister and Gaut, 2009; Zemach et al., 2010; Piya et al., 2017). The precise role of gene body methylation is more elusive, but it is proposed to maintain consistent expression levels by averting deposition of H2A histone variants in genic regions (Coleman-Derr and Zilberman, 2012).

There are many methods to study DNA methylation patterns at the genome level. The most widely used approach today is bisulfite sequencing. This method begins with genome-wide conversion of C to uracil, then to thymine after PCR amplification. Converted strands are sequenced using next generation technology. Finally, the sequenced reads are mapped to a reference genome, where methylated sites are identified by C to T conversion (Smith et al., 2009). Using bisulfite sequencing, the *Arabidopsis thaliana* methylome was the first to be reported. Studies found that CG

methylation was abundant in genic and TE regions, whereas non-CG methylation (CHG and CHH) was more prominent in TE-rich heterochromatin (Cokus et al., 2008; Lister et al., 2008).

Epigenetic inheritance relies on stable maintenance of methylation marks through mitotic and meiotic cell division. While genetically identical, epialleles are epigenetically distinct alleles stably inherited with corresponding phenotypes. The first epiallele study, for example, showed hypermethylation of the CYCLOIDEA promoter in *Linaria* resulted in natural flower symmetry variation (Cubas et al., 1999). It is widely accepted that DNA methylation may have evolved as a mechanism to cope with viral and TE genome invasion. Many naturally occurring epialleles arise in plant genomes by TE insertion near genic areas or through siRNAs-mediated DNA methylation. An epiallele at the colorless non-ripening (Cnr) locus for tomato may have been the result of CG hypermethylation at a nearby COPIA-like retrotransposon (Paszkowski and Grossniklaus, 2011). Similarly, recessive self-incompatibility alleles, S-alleles, in Brassicaceae are preferentially targeted for hypermethylation, giving rise to a mono-allelic gene expression pattern in heterozygotes (Kusaba et al., 2002).

3.3.1 Maintenance of DNA methylation

Models for DNA methylation maintenance are still being investigated; however, it is known that the mechanisms vary based on sequence context (symmetrical or asymmetrical) and genomic region (euchromatic or heterochromatic). Briefly, it can be divided into three steps (Figure 1.1)– (a) initiation of silencing pathways by recognition of dsRNA, (b) establishment of siRNA mediated DNA methylation, and (c) maintenance of already established methylation marks in the genome during replication. Initiation and establishment steps proceed according to PTGS and TGS as previously described. Methylation patterns in the CG context are maintained through replication by METHYLTRANSFERASE 1 (MET1). During DNA replication, chromatin binding proteins VARIANT IN METHYLATION (VIM) bind to hemimethylated parent strands and recruit MET1 (Kankel et al., 2003; Lister et al., 2008). Nucleosome remodeler

DECREASE in DNA METHYLATION 1 (DDM1) is hypothesized to aid MET1 in establishing methylation in heterochromatic and H1-rich regions (Zemach et al., 2013). Histone deacetylation is another common feature of heterochromatin. TE transcriptional silencing is lost in mutants lacking histone deacetylase 6 (HDA6) (Murfett et al., 2001; Lippman et al., 2003; Probst et al., 2004). Li et al., (2012) propose a model in which HDA6 directs deacetylation of H3 and H4 lysine residues on N-tails and recruit MET1 for establishment of CG methylation on DNA associated with these deacetylated histones (Li et al., 2012).

Non-CG methylation is maintained through a feedback loop between plant-specific CHROMOMETHYLTRANSFERASE 2 (CMT2) and CMT3 and histone methyltransferases. Dimethylation of histone H3 lysine 9 (H3K9me2) is established by histone methyltransferases-like Su(var)3–9 homolog4 (SUVH4) and its variants SUVH5 and SUVH6. These enzymes are directed towards methylated DNA by SET and RING finger-associated (SRA) domains, leading to the establishment of H3K9me2. Similarly, chromodomains and BAH domains of CMT2 and CMT3 interact with H3K9me2 and establish methylation in the CHG and CHH contexts, respectively (Bartee et al., 2001; Lindroth et al., 2001; Malagnac et al., 2002; Johnson et al., 2007; Bernatavichute et al., 2008; Du et al., 2012; Zemach et al., 2013; Stroud et al., 2014; Du et al., 2015). The information for symmetrical methylation contexts (CG and CHG) are retained on hemimethylated parental strands during semi-conservative DNA replication and therefore easily maintained. During maintenance of asymmetrical CHH methylation, one parent strand loses all methylation information during replication. CHH methylation marks are re-established via RDR2-RdDM pathway each replication cycle (Law and Jacobsen, 2010).

4. Epigenetic inheritance in response to stress

Stress memories are retained in somatic cells following exposure as a means for immunity towards future attacks; exemplified, for example, by plant defense priming and systemic acquired resistance (SAR) (Molinier et al., 2006). The primed state, or

stress memory is passed down the lineage through the progeny. However, inheritance of paternal effects is unstable because dividing cells often undergo reprogramming during differentiation and gametogenesis. Thus, parental inheritance is transgenerational, but only extends a few generations and gradually disappears from the genome after few cycles of meiosis. Although reprogramming in plants is not as stringent as in mammals, it still accounts for removal of epigenetic marks acquired from stress exposure by the parent (Heard and Martienssen, 2014). Small RNAs also play a crucial role in priming. Differential levels of small RNA template inherited from a primed parent may induce immunity in progeny; however, generational immunity diminishes once the stress stimuli is removed (Kachroo and Robin, 2013; Tricker, 2015). Importantly, histone modifications and cytoplasmically-transmitted small RNA-mediated are abolished during gametogenesis. Silencing of flower locus C (FLC) by histone methylation triggered flowering in adult plants through vernalization. However, this histone mark is reprogrammed and reset in germline cells (Sung and Amasino, 2004). Transgenerational stress immunity must be passed down through meiotically and mitotically stable epigenetic mechanisms. Cytosine methylation is the only epigenetic mark that can be altered in response to stimuli and stably inherited through generations.

Whole genome DNA methylation profiling of *drm1/drm2/cmt3* and *met1* mutants revealed that global reduction in DNA methylation results in enhanced resistance to bacterial infection (Dowen et al., 2012; Yu et al., 2013). The *drm1/drm2/cmt3* triple mutant developed a phenotype similar to a primed wild-type plant following infection with *Pseudomonas syringae*. DNA hypomethylation is mediated by SA-dependent defense signaling pathways, perhaps to activate expression of plant defense genes. Several studies have recently described DNA methylation as a controlling factor of compatibility between several phytopathogens and host plants (reviewed by Hewezi et al., 2017).

5. Epigenetics of host-plant parasitic nematode interactions

Epigenetic factors are fluent, driven by dynamic environmental cues. Phytopathogenic nematodes are key conductors of stress stimuli affecting host epigenomes. Arabidopsis mutants deprived of RdDM genes are more resistant to *Heterodera schachtii*, revealing that siRNAs and DNA methylation are important regulators of plant susceptibility to cyst nematode (Hewezi et al., 2008). This was recently validated by the accumulation of 24nt siRNA in conjunction with dramatic root methylome changes in response to *H. schachtii* (Hewezi et al., 2017). siRNAs aid in *de novo* DNA methylation by guiding DRM, which ultimately alter expression of several genes and TEs. In addition, stably inherited differential DNA methylation at certain loci modulate host susceptibility by regulating genes expressed during the initial phases of nematode infection. For example, rhg1 contains repeat units displaying cultivar-specific methylomes, perhaps under epigenetic control (Cook et al., 2014).

miRNAs also play an important role during PPN interactions, especially during feeding site establishment (Hewezi and Baum, 2015). Numerous miRNAs are found to be differentially expressed during nematode infection (Li et al., 2010; Zhang et al., 2011). However, many still await functional characterization. Deep sequencing of small RNA libraries from nematode-infected root tissues has provided clear insight into the dynamics of miRNA regulation during infection. Hewezi et al. (2008) first reported differentially expressed miRNA during various stages of *H. schachtii* infection in Arabidopsis, where more miRNA genes were downregulated than upregulated in early feeding site initiation. Since then, a similar trend has been observed in several plant–PPN interactions (Hewezi et al., 2008; Li et al., 2012; Cabrera et al., 2016). Many of these differentially expressed miRNAs target transcription factors to alter global expression (Hewezi et al., 2008; Hewezi et al., 2012; Li et al., 2012; Medina et al., 2017; Piya et al., 2017). For example, expression of GROWTH REGULATING FACTORS 1 (GRF1) and GRF3, which are involved in development and defense signaling, are inversely proportional to expression levels of their negative regulator, miR396 throughout plant-nematode parasitism (Hewezi et al., 2012). Similarly, miR390 impacts the expression levels of AUXIN REGULATING FACTORS (ARFs) in Arabidopsis roots

upon infection by the root knot nematode, *Meloidogyne javanica* (Cabrera et al., 2016). Further, Zhao et al. (2015) provided a functional role of miR319b and its target TCP4 in modulating early JA-mediated defenses in tomato. Transgenic tomato overexpressing miR319b were more susceptible to root knot, whereas overexpression of TCP4 in mutants increased resistance to infection (Zhao et al., 2015). MYB family transcription factors are also targeted by miRNAs to regulate cellular processes, cell wall biosynthesis, and stress responses (Cominelli et al., 2005; McCarthy et al., 2010; Oh et al., 2011). MYB83 and MYB33 are regulated by miR858 and miRNA159, respectively, and demonstrate important roles in determining nematode host susceptibility (Medina et al., 2017; Piya et al., 2017). One case of NITROGEN LIMITATION ADAPTATION (NLA) gene-miRNA regulation has been found in the Arabidopsis x *H. schachtii* pathosystem, where miR827 controls basal defense response (Hewezi et al., 2016).

There is ample evidence indicating PPNs exploit various epigenetic mechanisms to successfully establish and perpetuate disease. Nevertheless, the interaction of these epigenetic modifications for host transcriptional reprogramming is unknown during infection. Therefore, there is a strong interest in determining the extent to which inducible and heritable DNA methylation shapes compatible and incompatible interactions between SCN and soybean. The impact of DNA methylation on miRNA activity must also be assessed during SCN infection for a more comprehensive understanding. These topics are the primary focus of this thesis and are addressed in chapters 2, 3, and 4.

6. Dissertation organization

This dissertation is organized in five chapters to summarize the main objectives of the research project.

Chapter 1: General introduction provides information regarding soybean cyst nematode disease and advances made so far in search of resistance in soybean. Also, provides information about epigenetic modifications and their role in regulating stress response in plants especially against plant parasitic nematodes.

Chapter 2: Whole genome DNA methylation changes and their role in regulating gene expression during SCN infection were reported by studying SCN susceptible soybean cultivar Williams 82.

Chapter 3: Differences in DNA methylation changes between compatible and incompatible SCN interactions were deduced by comparing two highly homozygous near-isogenic lines differing at soybean cyst nematode resistance locus, Rhg4.

Chapter 4: Differential methylation in microRNA genes during SCN infection were identified for compatible and incompatible interactions and impact of their target genes on soybean susceptibility was assessed.

Chapter 5: General conclusions and future directions for further research were discussed in this chapter.

References

Adee EA, Johnson ML, Niblack TL (2008) Effect of soybean cultivars moderately resistant to soybean cyst nematode on SCN populations and yield. Plant Health Progress doi **10**

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

Anand S, Rao-Arelli A (1989) Genetic analyses of soybean genotypes resistant to soybean cyst nematode race 5. Crop science **29:** 1181-1184

Ankeny RA (2001) The natural history of *Caenorhabditis elegans* research. Nature reviews. Genetics **2:** 474

Axtell MJ (2013) Classification and comparison of small RNAs from plants. Annual review of plant biology **64:** 137-159

Bannister AJ, Kouzarides T (2011) Regulation of chromatin by histone modifications. Cell research **21:** 381

Bartee L, Malagnac F, Bender J (2001) Arabidopsis cmt3 chromomethylase mutations block non-CG methylation and silencing of an endogenous gene. Genes & development **15:** 1753-1758

Bergmüller E, Gehrig PM, Gruissem W (2007) Characterization of post-translational modifications of histone H2B-variants isolated from *Arabidopsis thaliana*. Journal of proteome research **6:** 3655-3668

Bernatavichute YV, Zhang X, Cokus S, Pellegrini M, Jacobsen SE (2008) Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in *Arabidopsis thaliana*. PloS one **3:** e3156

Borges F, Martienssen RA (2015) The expanding world of small RNAs in plants. Nature reviews. Molecular cell biology **16:** 727

Cabrera J, Barcala M, García A, Rio-Machín A, Medina C, Jaubert-Possamai S, Favery B, Maizel A, Ruiz-Ferrer V, Fenoll C (2016) Differentially

expressed small RNAs in Arabidopsis galls formed by *Meloidogyne javanica*: a functional role for miR390 and its TAS3-derived tasiRNAs. New Phytologist **209:** 1625-1640

Cao X, Jacobsen SE (2002) Role of the Arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. Current Biology **12:** 1138-1144

Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE (2008) Shotgun bisulfite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature **452:** 215

Coleman-Derr D, Zilberman D (2012) Deposition of histone variant H2A. Z within gene bodies regulates responsive genes. PLoS genetics **8:** e1002988

Cominelli E, Galbiati M, Vavasseur A, Conti L, Sala T, Vuylsteke M, Leonhardt N, Dellaporta SL, Tonelli C (2005) A guard-cell-specific MYB transcription factor regulates stomatal movements and plant drought tolerance. Current biology **15:** 1196-1200

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

Cook DE, Bayless AM, Wang K, Guo X, Song Q, Jiang J, Bent AF (2014) Distinct copy number, coding sequence, and locus methylation patterns underlie Rhg1-mediated soybean resistance to soybean cyst nematode. Plant physiology **165:** 630-647

Cook DE, Lee TG, Guo X, Melito S, Wang K, Bayless AM, Wang J, Hughes TJ, Willis DK, Clemente TE (2012) Copy number variation of multiple genes at Rhg1 mediates nematode resistance in soybean. Science **338:** 1206-1209

Cubas P, Vincent C, Coen E (1999) An epigenetic mutation responsible for natural variation in floral symmetry. Nature **401:** 157

Cuerda-Gil D, Slotkin RK (2016) Non-canonical RNA-directed DNA methylation. Nature plants **2:** 16163

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

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

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

Du J, Johnson LM, Jacobsen SE, Patel DJ (2015) DNA methylation pathways and their crosstalk with histone methylation. Nature reviews. Molecular cell biology **16:** 519

Du J, Zhong X, Bernatavichute YV, Stroud H, Feng S, Caro E, Vashisht AA, Terragni J, Chin HG, Tu A (2012) Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. Cell **151:** 167-180

Ferris H (2010) Contribution of nematodes to the structure and function of the soil food web. Journal of nematology **42:** 63

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

Heard E, Martienssen RA (2014) Transgenerational epigenetic inheritance: myths and mechanisms. Cell **157:** 95-109

Hewezi T, Baum TJ (2013) Manipulation of plant cells by cyst and root-knot nematode effectors. Molecular Plant-Microbe Interactions **26:** 9-16

Hewezi T, Baum TJ (2015) Chapter Nine-Gene Silencing in Nematode Feeding Sites. Advances in botanical research **73:** 221-239

Hewezi T, Howe P, Maier TR, Baum TJ (2008) Arabidopsis small RNAs and their targets during cyst nematode parasitism. Molecular Plant-Microbe Interactions **21:** 1622-1634

Hewezi T, Lane T, Piya S, Rambani A, Rice JH, Staton M (2017) Cyst nematode parasitism induces dynamic changes in the root epigenome. Plant Physiology**:** pp. 01948.02016

Hewezi T, Maier TR, Nettleton D, Baum TJ (2012) The Arabidopsis microRNA396-GRF1/GRF3 regulatory module acts as a developmental regulator in the reprogramming of root cells during cyst nematode infection. Plant physiology **159:** 321-335

Hewezi T, Pantalone V, Bennett M, Stewart CN, Burch-Smith TM (2017a) Phytopathogen-induced changes to plant methylomes. Plant Cell Reports**:** 1-7

Hewezi T, Piya S, Qi M, Balasubramaniam M, Rice JH, Baum TJ (2016) Arabidopsis miR827 mediates post-transcriptional gene silencing of its ubiquitin E3 ligase target gene in the syncytium of the cyst nematode *Heterodera schachtii* to enhance susceptibility. The Plant Journal **88:** 179-192

Hollister JD, Gaut BS (2009) Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome research **19:** 1419-1428

Hymowitz T, Harlan JR (1983) Introduction of soybean to North America by Samuel Bowen in 1765. Economic Botany **37:** 371-379

Inagaki H, TsuTsuMI M (1971) Survival of the Soybean Cyst Nematode, Heterodera glycines ICHINOHE (Tylenchida: Heteroderidae) under Certain String Conditions. Applied entomology and zoology **6:** 156-162

Ithal N, Mitchum MG (2011) Laser capture microdissection of nematode feeding cells. Plant Immunity: Methods and Protocols**:** 227-240

Jackson JP, Lindroth AM, Cao X, Jacobsen SE (2002) Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature **416:** 556-560

Johnson LM, Bostick M, Zhang X, Kraft E, Henderson I, Callis J, Jacobsen SE (2007) The SRA methyl-cytosine-binding domain links DNA and histone methylation. Current Biology **17:** 379-384

Jones M (1981) Host cell responses to endoparasitic nematode attack: structure and function of giant cells and syncytia. Annals of applied Biology **97:** 353-372

Jones M, Tan J, Fosu-Nyarko J (2014) Candidate secreted proteins and effectors of Pratylenchus SPP.

Kachroo A, Robin GP (2013) Systemic signaling during plant defense. Current Opinion in Plant Biology **16:** 527-533

Kandoth PK, Mitchum MG (2013) War of the worms: how plants fight underground attacks. Current opinion in plant biology **16:** 457-463

Kankel MW, Ramsey DE, Stokes TL, Flowers SK, Haag JR, Jeddeloh JA, Riddle NC, Verbsky ML, Richards EJ (2003) Arabidopsis MET1 cytosine methyltransferase mutants. Genetics **163:** 1109-1122

Khan R, Alkharouf N, Beard H, MacDonald M, Chouikha I, Meyer S, Grefenstette J, Knap H, Matthews B (2004) Microarray analysis of gene expression in soybean roots susceptible to the soybean cyst nematode two days post invasion. Journal of nematology **36:** 241

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

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

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

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

Klink VP, Overall CC, Alkharouf NW, MacDonald MH, Matthews BF (2010) Microarray detection call methodology as a means to identify and compare transcripts expressed within syncytial cells from soybean (*Glycine max*) roots undergoing resistant and susceptible reactions to the soybean cyst nematode (*Heterodera glycines*). Journal of BioMed Research **2010**

Koenning SR, Wrather JA (2010) Suppression of soybean yield potential in the continental United States by plant diseases from 2006 to 2009. Plant Health Progress **10**

Kusaba M, Tung C-W, Nasrallah ME, Nasrallah JB (2002) Monoallelic expression and dominance interactions in anthers of self-incompatible *Arabidopsis lyrata*. Plant Physiology **128:** 17-20

Lamers P, Rosillo-Calle F, Pelkmans L, Hamelinck C (2014) Developments in international liquid biofuel trade. *In* International Bioenergy Trade. Springer, pp 17-40

Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature reviews. Genetics **11:** 204

Lee TG, Kumar I, Diers BW, Hudson ME (2015) Evolution and selection of Rhg1, a copy-number variant nematode-resistance locus. Molecular ecology **24:** 1774-1791

Li X, Qian W, Zhao Y, Wang C, Shen J, Zhu J-K, Gong Z (2012) Antisilencing role of the RNA-directed DNA methylation pathway and a histone acetyltransferase in Arabidopsis. Proceedings of the National Academy of Sciences **109:** 11425-11430

Li X, Wang X, Zhang S, Liu D, Duan Y, Dong W (2012) Identification of soybean microRNAs involved in soybean cyst nematode infection by deep sequencing. PloS one **7:** e39650

Li Y, Zhang Q, Zhang J, Wu L, Qi Y, Zhou J-M (2010) Identification of microRNAs involved in pathogen-associated molecular pattern-triggered plant innate immunity. Plant physiology **152:** 2222-2231

Lindroth AM, Cao X, Jackson JP, Zilberman D, McCallum CM, Henikoff S, Jacobsen SE (2001) Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. Science **292:** 2077-2080

Lippman Z, May B, Yordan C, Singer T, Martienssen R (2003) Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. PLoS biology **1:** e67

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell **133:** 523-536

Liu S, Kandoth PK, Warren SD, Yeckel G, Heinz R, Alden J, Yang C, Jamai A, El-Mellouki T, Juvale PS (2012) A soybean cyst nematode resistance gene points to a new mechanism of plant resistance to pathogens. Nature **492:** 256

Liu X, Liu S, Jamai A, Bendahmane A, Lightfoot DA, Mitchum MG, Meksem K (2011) Soybean cyst nematode resistance in soybean is independent of the Rhg4 locus LRR-RLK gene. Functional & integrative genomics **11:** 539-549

Luger K, Dechassa ML, Tremethick DJ (2012) New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? Nature reviews Molecular cell biology **13:** 436-447

Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 angstrom resolution. Nature **389:** 251

Mahrez W, Arellano MST, Moreno-Romero J, Nakamura M, Shu H, Nanni P, Köhler C, Gruissem W, Hennig L (2016) H3K36ac is an evolutionary conserved plant histone modification that marks active genes. Plant physiology**:** pp. 01744.02015

Malagnac F, Bartee L, Bender J (2002) An Arabidopsis SET domain protein required for maintenance but not establishment of DNA methylation. The EMBO journal **21:** 6842-6852

Mazarei M, Liu W, Al-Ahmad H, Arelli PR, Pantalone VR, Stewart CN (2011) Gene expression profiling of resistant and susceptible soybean lines infected with soybean cyst nematode. Theoretical and Applied Genetics **123:** 1193

McCarthy RL, Zhong R, Fowler S, Lyskowski D, Piyasena H, Carleton K, Spicer C, Ye Z-H (2010) The poplar MYB transcription factors, PtrMYB3 and PtrMYB20, are involved in the regulation of secondary wall biosynthesis. Plant and Cell Physiology **51:** 1084-1090

McKinley KL, Cheeseman IM (2016) The molecular basis for centromere identity and function. Nature reviews. Molecular cell biology **17:** 16

Medina C, Rocha M, Magliano M, Ratpopoulo A, Revel B, Marteu N, Magnone V, Lebrigand K, Cabrera J, Barcala M (2017) Characterization of microRNAs from Arabidopsis galls highlights a role for miR159 in the plant

response to the root-knot nematode *Meloidogyne incognita*. New Phytologist **216.3:**882-896

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

Melito S, Heuberger AL, Cook D, Diers BW, MacGuidwin AE, Bent AF (2010) A nematode demographics assay in transgenic roots reveals no significant impacts of the Rhg1 locus LRR-Kinase on soybean cyst nematode resistance. BMC plant biology **10:** 104

Molinier J, Ries G, Zipfel C, Hohn B (2006) Transgeneration memory of stress in plants. Nature **442:** 1046

Murfett J, Wang X-J, Hagen G, Guilfoyle TJ (2001) Identification of Arabidopsis histone deacetylase HDA6 mutants that affect transgene expression. The Plant Cell **13:** 1047-1061

Niblack T, Lambert K, Tylka G (2006) A model plant pathogen from the kingdom animalia: *Heterodera glycines*, the soybean cyst nematode. Annu. Rev. Phytopathol. **44:** 283-303

Nicol JM, Turner SJ, Coyne D, Den Nijs L, Hockland S, Maafi ZT (2011) Current nematode threats to world agriculture. *In* Genomics and molecular genetics of plant-nematode interactions. Springer, pp 21-43

Oh JE, Kwon Y, Kim JH, Noh H, Hong S-W, Lee H (2011) A dual role for MYB60 in stomatal regulation and root growth of *Arabidopsis thaliana* under drought stress. Plant molecular biology **77:** 91-103

Onodera Y, Haag JR, Ream T, Nunes PC, Pontes O, Pikaard CS (2005) Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. Cell **120:** 613-622

Paszkowski J, Grossniklaus U (2011) Selected aspects of transgenerational epigenetic inheritance and resetting in plants. Current opinion in plant biology **14:** 195-203

Pikaard CS, Haag JR, Ream T, Wierzbicki AT (2008) Roles of RNA polymerase IV in gene silencing. Trends in plant science **13:** 390-397

Piya S, Bennett M, Rambani A, Hewezi T (2017a) Transcriptional activity of transposable elements may contribute to gene expression changes in the syncytium formed by cyst nematode in arabidopsis roots. Plant Signaling & Behavior **12.9:** e1362521

Piya S, Kihm C, Rice JH, Baum TJ, Hewezi T (2017) Cooperative regulatory functions of miR858 and MYB83 during cyst nematode parasitism. Plant Physiology **174:** 1897-1912

Probst AV, Fagard M, Proux F, Mourrain P, Boutet S, Earley K, Lawrence RJ, Pikaard CS, Murfett J, Furner I (2004) Arabidopsis histone deacetylase HDA6 is required for maintenance of transcriptional gene silencing and determines nuclear organization of rDNA repeats. The Plant Cell **16:** 1021-1034

Rambani A, Rice JH, Liu J, Lane T, Ranjan P, Mazarei M, Pantalone V, Stewart CN, Staton M, Hewezi T (2015) The methylome of soybean roots during the compatible interaction with the soybean cyst nematode. Plant physiology **168:** 1364-1377

Rao-Arelli A, Anand S (1988) Genetic relationships among soybean plant introductions for resistance to race 3 of soybean cyst nematode. Crop science **28:** 650-652

Roudier F, Ahmed I, Bérard C, Sarazin A, Mary-Huard T, Cortijo S, Bouyer D, Caillieux E, Duvernois-Berthet E, Al-Shikhley L (2011) Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. The EMBO journal **30:** 1928-1938

Rowley MJ, Rothi MH, Böhmdorfer G, Kuciński J, Wierzbicki AT (2017) Long-range control of gene expression via RNA-directed DNA methylation. PLoS genetics **13:** e1006749

Sengupta P, Samuel AD (2009) *Caenorhabditis elegans*: a model system for systems neuroscience. Current opinion in neurobiology **19:** 637-643

Sequeira-Mendes J, Aragüez I, Peiró R, Mendez-Giraldez R, Zhang X, Jacobsen SE, Bastolla U, Gutierrez C (2014) The functional topography of the Arabidopsis genome is organized in a reduced number of linear motifs of chromatin states. The Plant Cell **26:** 2351-2366

Shu H, Nakamura M, Siretskiy A, Borghi L, Moraes I, Wildhaber T, Gruissem W, Hennig L (2014) Arabidopsis replacement histone variant H3. 3 occupies promoters of regulated genes. Genome biology **15:** R62

Sifri CD, Begun J, Ausubel FM (2005) The worm has turned–microbial virulence modeled in *Caenorhabditis elegans*. Trends in microbiology **13:** 119-127

Slotkin RK, Freeling M, Lisch D (2005) Heritable transposon silencing initiated by a naturally occurring transposon inverted duplication. Nature genetics **37:** 641

Smith ZD, Gu H, Bock C, Gnirke A, Meissner A (2009) High-throughput bisulfite sequencing in mammalian genomes. Methods **48:** 226-232

Steiner FA, Henikoff S (2015) Diversity in the organization of centromeric chromatin. Current opinion in genetics & development **31:** 28-35

Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, Patel DJ, Jacobsen SE (2014) Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. Nature structural & molecular biology **21:** 64-72

Stroud H, Otero S, Desvoyes B, Ramírez-Parra E, Jacobsen SE, Gutierrez C (2012) Genome-wide analysis of histone H3. 1 and H3. 3 variants in *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences **109:** 5370-5375

Struhl K (1998) Histone acetylation and transcriptional regulatory mechanisms. Genes & development **12:** 599-606

Sung S, Amasino RM (2004) Vernalization in *Arabidopsis thaliana* is mediated by the PHD finger protein VIN3. Nature **427:** 159-164

Talbert PB, Henikoff S (2010) Histone variants--ancient wrap artists of the epigenome. Nature reviews. Molecular cell biology **11:** 264

Talbert PB, Henikoff S (2014) Environmental responses mediated by histone variants. Trends in cell biology **24:** 642-650

Talbert PB, Henikoff S (2016) Histone variants on the move: substrates for chromatin dynamics. Nature Reviews Molecular Cell Biology

Thomas JO (1999) Histone H1: location and role. Current opinion in cell biology **11:** 312-317

Triantaphyllou A (1973) Environmental sex differentiation of nematodes in relation to pest management. Annual review of phytopathology **11:** 441-462

Triantaphyllou A, Hirschmann H (1962) Oogenesis and Mode of Reproduction in the Soybean Cyst Nematode, *Heterodera glycines*. Nematologica **7:** 235-241

Tricker PJ (2015) Transgenerational inheritance or resetting of stress-induced epigenetic modifications: two sides of the same coin. Frontiers in plant science **6**

Tytgat T, De Meutter J, Gheysen G, Coomans A (2000) Sedentary endoparasitic nematodes as a model for other plant parasitic nematodes. Nematology **2:** 113-121

Wang J, Niblack T, Tremain J, Wiebold W, Tylka G, Marett C, Noel G, Myers O, Schmidt M (2003) Soybean cyst nematode reduces soybean yield without causing obvious aboveground symptoms. Plant Disease **87:** 623-628

Weiberg A, Wang M, Bellinger M, Jin H (2014) Small RNAs: a new paradigm in plant-microbe interactions. Annual review of phytopathology **52**

Wierzbicki AT, Haag JR, Pikaard CS (2008) Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. Cell **135:** 635-648

Wierzbicki AT, Ream TS, Haag JR, Pikaard CS (2009) RNA polymerase V transcription guides ARGONAUTE4 to chromatin. Nature genetics **41:** 630-634

Wollmann H, Holec S, Alden K, Clarke ND, Jacques P-E, Berger F (2012) Dynamic deposition of histone variant H3. 3 accompanies developmental remodeling of the Arabidopsis transcriptome. PLoS Genetics **8:** e1002658

Wrather J, Koenning S (2009) Effects of diseases on soybean yields in the United States 1996 to 2007. Virus **2:** 9

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

Wyss U, Zunke U (1986) Observations on the behaviour of second stage juveniles of Hetero inside host roots. Revue Nematol **9:** 153-165

Yelagandula R, Stroud H, Holec S, Zhou K, Feng S, Zhong X, Muthurajan UM, Nie X, Kawashima T, Groth M (2014) The histone variant H2A. W defines heterochromatin and promotes chromatin condensation in Arabidopsis. Cell **158:** 98-109

Yu A, Lepère G, Jay F, Wang J, Bapaume L, Wang Y, Abraham A-L, Penterman J, Fischer RL, Voinnet O (2013) Dynamics and biological

relevance of DNA demethylation in Arabidopsis antibacterial defense. Proceedings of the National Academy of Sciences **110:** 2389-2394

Zemach A, Kim MY, Hsieh P-H, Coleman-Derr D, Eshed-Williams L, Thao K, Harmer SL, Zilberman D (2013) The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell **153:** 193-205

Zemach A, McDaniel IE, Silva P, Zilberman D (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science **328:** 916-919

Zhang H, Zhu J-K (2011) RNA-directed DNA methylation. Current opinion in plant biology **14:** 142-147

Zhang K, Sridhar VV, Zhu J, Kapoor A, Zhu J-K (2007) Distinctive core histone post-translational modification patterns in *Arabidopsis thaliana*. PloS one **2:** e1210

Zhang W, Gao S, Zhou X, Chellappan P, Chen Z, Zhou X, Zhang X, Fromuth N, Coutino G, Coffey M (2011) Bacteria-responsive microRNAs regulate plant innate immunity by modulating plant hormone networks. Plant molecular biology **75:** 93-105

Zhao W, Li Z, Fan J, Hu C, Yang R, Qi X, Chen H, Zhao F, Wang S (2015) Identification of jasmonic acid-associated microRNAs and characterization of the regulatory roles of the miR319/TCP4 module under root-knot nematode stress in tomato. Journal of experimental botany **66:** 4653-4667

Zilberman D, Cao X, Jacobsen SE (2003) ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. Science **299:** 716-719

Zilberman D, Coleman-Derr D, Ballinger T, Henikoff S (2008) Histone H2A. Z and DNA methylation are mutually antagonistic chromatin marks. Nature **456:** 125

Appendix

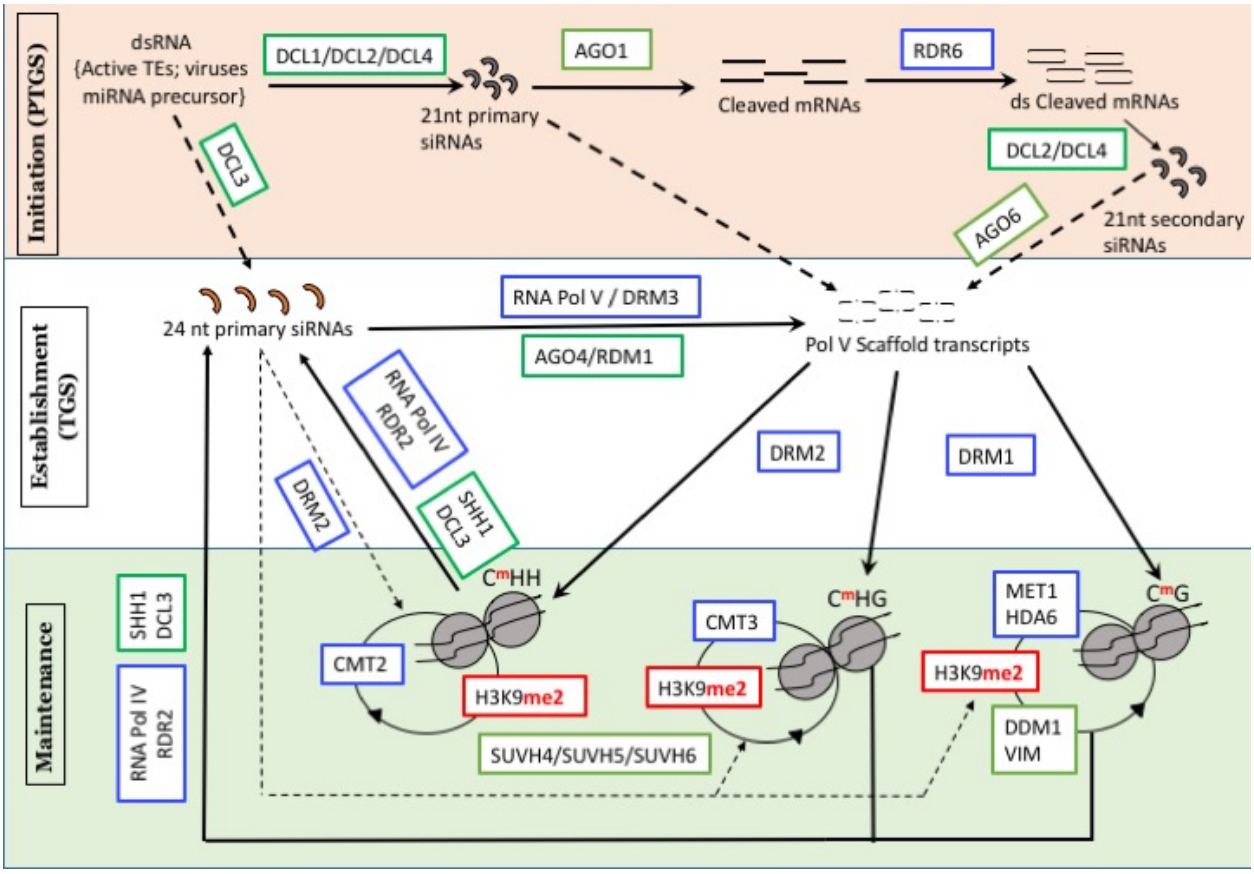

Figure 1.1 Model of pathways involved in initiation, establishment and maintenance of DNA methylation in plants.

Chapter 2: The methylome of soybean roots during the compatible interaction with the soybean cyst nematode, *Heterodera glycines*

Title: The methylome of soybean roots during the compatible interaction with the soybean cyst nematode, *Heterodera glycines*

Aditi Rambani[1], J. Hollis Rice[1], Jinyi Liu[1], Thomas Lane[2], Priya Ranjan[1], Mitra Mazarei[1], Vince Pantalone[1], C. Neal Stewart Jr.[1], Meg Staton[2] and Tarek Hewezi[1]*
[1]Department of Plant Sciences, University of Tennessee, Knoxville, Tennessee 37996.
[2]Department of Entomology, Plant Pathology and Nematology, University of Tennessee, Knoxville, Tennessee 37996.

Published in *Plant physiology*, *168*(4), pp.1364-1377.

I helped design experiments for this project. I prepared bisulfite sequencing libraries and contributed towards analysis of methylome data. I contributed towards writing portion of the manuscript describing the methods I performed.

Abstract

Soybean cyst nematode (SCN, *Heterodera glycines*) induces the formation of a multinucleated feeding site, or syncytium, whose etiology includes massive gene expression changes. Nevertheless, the genetic networks underlying gene expression control in the syncytium are poorly understood. DNA methylation is a critical epigenetic mark that plays a key role in regulating gene expression. To determine the extent to which DNA methylation is altered in soybean roots during the susceptible interaction with SCN, we generated whole-genome cytosine methylation maps at single nucleotide resolution. The methylome analysis revealed that SCN induces hypo-methylation to a much higher extent than hyper-methylation. We identified 2,465 differentially hyper-methylated regions and 4,692 hypo-methylated regions in the infected roots compared with the non-infected control. In addition, a total number of 703 and 1346 unique genes were identified as overlapping with hyper- or hypo-methylated regions, respectively. The differential methylation in genes apparently occurs independently of gene size and GC content but exhibits strong preference for recently duplicated paralogs. Furthermore, a set of 278 genes was identified as specifically syncytium differentially methylated genes. Of these, we found genes associated with epigenetic regulation, phytohormone signaling, cell wall architecture, signal transduction and ubiquitination. This study provides new evidence that differential methylation is part of the regulatory mechanisms controlling gene expression changes in the nematode-induced syncytium.

1. Introduction

Soybean cyst nematode (SCN, *Heterodera glycines*) is a sedentary parasite of soybean roots and is considered to be the most serious pathogen problem in soybean production worldwide (Koenning and Wrather, 2010). SCN undergoes a first molt inside the egg to develop from a first-stage (J1) to a second-stage juvenile (J2) before hatching. Hatched, pre-parasitic J2s then penetrate into soybean roots by breaching epidermal cell walls with their protrusible stylet, a hollow oral spear, aided by cell wall-digesting enzymes and other effector proteins (Hewezi and Baum, 2013). The parasitic J2s migrate to the vicinity of the vascular tissues where the nematodes become sedentary and begin feeding. In order to maintain the sedentary lifestyle, SCN induces the formation of a multinucleated feeding site in roots, the syncytium, whose etiology includes considerable cell-to-cell fusion, in addition to dramatic cytoplasmic and nuclear modifications (Williamson and Hussey, 1996). Molecular and genetic studies in the last few years have provided new insights into the molecular mechanisms associated with SCN parasitism. Global gene expression profiling using microarrays has been extensively studied during plant-nematode interactions. Initially these studies used the whole soybean root system to quantify transcriptional changes associated with the compatible interaction (Khan et al., 2004; Alkharouf et al., 2006). More comprehensive gene expression studies were conducted on syncytial cells using laser capture microdissection (LCM) coupled with microarray analysis (Ithal et al., 2007; Klink et al., 2007; Klink et al., 2009; Klink et al., 2010; Kandoth et al., 2011). These cell-specific gene expression analyses yielded lists of thousands of syncytial differentially expressed genes and provided new information about the underlying molecular events occurring during syncytium formation and function. However, the genetic networks underlying gene expression regulation in nematode-infected roots and particularly in the nematode-induced feeding sites are poorly understood. Because epigenetic modifications function in concert with genetic mechanisms to regulate gene expression in normal cells and are often dysregulated in infected cells, these modifications may contribute significantly to the regulation of the transcriptional activity of syncytial cells.

Epigenetics include biochemical modifications of DNA and associated proteins that regulate gene expression and chromosome structure and function, without changing

DNA nucleotide sequences. DNA methylation, the most common epigenetic modification, is the addition or removal of a methyl group (CH_3), mostly where cytosine bases occur repeatedly. In plants, DNA methylation occurs in symmetric (CpG and CHG) and asymmetric (CHH) contexts where H refers to any nucleotide but G. The CpG and CHG patterns are symmetric across the two DNA strands, which are believed to be important for the maintenance of methylation at these sites following DNA replication. DNA cytosine methylation controls gene expression networks and hence plays essential roles in different aspects of plant growth, development, and response to biotic stress (Alvarez et al., 2010; Zhang et al., 2010; He et al., 2011; Dowen et al., 2012). In plants, *de novo* DNA methylation in all sequence contexts is mainly mediated through the activity of DOMAINS REARRANGED METHYLTRANSFERASEs (DRMs), homologs of the mammalian DNA methyltransferase 3 (DNMT3) enzymes (Cao and Jacobsen, 2002; Cao et al., 2003). The DRMs are guided to specific DNA sequences through the RNA-directed DNA methylation (RdDM) pathway, which involves the generation of ~23- to 24-nt siRNAs by RNA-dependent RNA polymerase 2 (RDR2) and DICER-Like 3 (DCL3) (Henderson and Jacobsen, 2007). These siRNAs are loaded into a complex containing the argonaute (AGO) proteins AGO4 and AGO6 to guide DRMs to target loci by a currently unknown mechanism (Matzke et al., 2009).

While DNA methylation has been initially reported to control various developmental processes in plants (He et al., 2011), recent studies revealed that this silencing pathway plays a key role in modulating plant defense responses during biotrophic interactions. Compelling evidence of dynamic changes in DNA methylation in response to infection by the bacterial pathogen *Pseudomonas syringae* pv. *tomato* DC3000 (*Pst*) has been recently reported (Dowen et al., 2012). Using deep sequencing of bisulfite-treated DNA, Dowen et al. (2012) found that differentially methylated regions were preferentially associated with genes involved in defense response, and that hypomethylation in differentially methylated regions were frequently accompanied by activation of the proximal genes, specifically those with defense response function. Similarly, another recent study indicated that DNA demethylation restricted the multiplication and

vascular propagation of the *Pst* and, consequently, some immune response genes were repressed by DNA methylation (Yu et al., 2013).

The role of DNA methylation in regulating the plant immune system was further supported by the finding that mutant lines entirely defective in maintenance of CpG methylation or non-CpG methylation were resistance to *Pst* infection (Dowen et al., 2012). Likewise, mutants partially impaired in CpG methylation (decrease in DNA methylation1, *ddm1*) or non-CpG methylation (RNA-dependent RNA polymerase mutants *rdr1, rdr2* and *rdr6*, and DICER-like triple mutant *dcl2/3/4*) showed modest increases in *Pst* resistance (Dowen et al., 2012). These data are consistent with our previously published results in which we found that several Arabidopsis *dcl* and *rdr* mutants were more resistant to the beet cyst nematode *H. schachtii* compared to wild type (Hewezi et al., 2008). Chemical demethylation of the silenced resistance *Xa21G* gene in rice reestablished its resistance function against *Xanthomomonas oryzae* (Akimoto et al., 2007). Similarly, induced DNA hypomethylation at the NBS–LRR gene clusters by the tobacco mosaic virus was associated with increased genomic rearrangements at these genomic loci (Boyko et al., 2007). The expression difference between the resistant alleles of the *Medicago truncatula REP1* gene, which confers resistance against the powdery mildew disease caused by the biotrophic fungus *Erysiphe pisi,* was found to be correlated with the methylation status of the promoter regions (Yang et al., 2013). In soybean, differential hyper-methylation patterns at the genomic regions that contain multiple copies of the SCN resistance gene *Rhg1* have been recently identified (Cook et al., 2014). Collectively these results indicate that DNA methylation plays a crucial role in regulating the immune system in response to infection by plant pathogens including cyst nematodes.

In this study, we used the Methyl-Seq method to generate high-coverage genome-wide DNA methylation maps encompassing the methylation status of single cytosine bases throughout the genome of soybean roots during the susceptible interactions with SCN. A high number of differentially methylated regions and overlapping protein-coding genes were identified. In addition, we found a significant portion of the differentially methylated genes was among genes reported to change expression in the soybean

syncytium, pointing to a novel role of SCN-induced DNA methylation in regulating gene expression changes during parasitism.

2. Results

2.1 Genome-wide DNA methylation profiling of soybean roots during the susceptible interaction with the soybean cyst nematode, *Heterodera glycines*

In order to profile the DNA methylation patterns at single nucleotide resolution during the susceptible interaction with SCN, we construction six whole genome bisulfite-treated DNA libraries. In these experiments, we inoculated soybean cultivar Williams 82 with SCN (HG type 0 or race 3) and root tissues were collected at 5 day post inoculation (dpi) from both infected and non-infected soybean roots from three independent experiments. Two libraries (infected and control) from each experiment for a total of six libraries were generated and sequenced using the Illumina HiSeq platform. The sequencing data were grouped into six different files and a total number of 175 million 100 bp reads for SCN-infected samples and 182 million reads for non-infected control were obtained. Because of the paleopolyploid nature of soybean genome, a significant portion of these reads were mapped to more than one genomic location and thus these reads were discarded. After quality filtering a total of about 60 million reads for each of the SCN-infected and non-infected samples were of high quality and uniquely mapped to soybean genome (Wm82.a2.v1) (Schmutz et al., 2010). These high quality sequence reads represented about 10X genome coverage, a depth greater than previously reported in Arabidopsis and soybean (Dowen et al., 2012; Schmitz et al., 2013). Bisulfite conversion efficiency was higher than 99% as determined using the non-methylated lambda phage genome. The percentage of methylated cytosines in CpG, CHG and CHH contexts were very similar across the three biological replicates for SCN-infected and non-infected control. The SCN-infected samples had an average of 78.6%, 56.4% and 5.0% methylation in overall cytosines occurring in CpG, CHG or CHH contexts,

respectively. Similarly, the control samples had an average of 81%, 59.3% and 5.5% methylation in overall cytosines occurring in CpG, CHG or CHH contexts, respectively. These data indicated that overall average methylation levels were very similar between the SCN-infected and control samples.

2.2 Identification of differentially methylated regions and associated genes

Because DNA methylation tends to cluster in specific regions of the genome (Zhang et al., 2006; Vaughn et al., 2007; Becker et al., 2011), and differences in single cytosines may not be of functional importance, a 200 bp non-overlapping sliding window approach was used across the soybean genome to identify differentially methylated regions and overlapping genes. Differentially hyper- and hypo-methylated regions between SCN-infected and control in CpG, CHG and CHH contexts were identified using *P*-value <0.01 and a percent methylation difference larger than 25%. Interestingly, in CpG context 718 hyper-methylated regions and 1408 hypo-methylated regions were identified in the infected roots compared with the non-infected control (Figure 2.1A). Similarly, in CHG context, 1142 hyper-methylated regions and 2074 hypo-methylated regions were identified (Figure 2.1A), whereas in CHH context 605 hyper-methylated regions and 1210 hypo-methylated regions were identified (Figure 2.1A). These results indicate that SCN induces hypo-methylation to much higher extent than hyper-methylation.

We assigned each differentially methylated regions to proximal protein-coding genes based on the genomic coordinates. We found 60% (428 genes), 16% (180 genes), and 20% (120 genes) of the hyper-methylated regions in CpG, CHG and CHH contexts, respectively, overlapped with protein-coding genes (Figure 2.1B). Similarly, 58% (817 genes), 17% (350 genes), and 23% (282 genes) of the hypo-methylated regions in CpG, CHG and CHH contexts, respectively overlapped with protein-coding genes (Figure 2.1C). As a result, a total number of 703 and 1346 unique genes were identified as hyper- and hypo-methylated, respectively. We next examined whether various methylation

contexts occur in individual genes. Interestingly, we identified 25 genes that were hyper-methylated in more than one context. Six of these genes were found to be hyper methylated in both CHH and CHG contexts ($\chi 2$ =34.862, P = 1.30E-07), 6 genes in CHH and CpG ($\chi 2$ =9.275, P = 0.026), and 13 genes in CpG and CHG contexts ($\chi 2$ =37.786, P =3.138E-08), (Figure 2.2B). Also, we identified 12 genes that were hypo-methylated in both CHH and CHG contexts ($\chi 2$ =20.059, P = 0.00017), 15 genes in CHH and CpG ($\chi 2$ =5.392, P = 0.145), and 78 genes in CpG and CHG contexts ($\chi 2$ = 457.025, P = 9.797E-99), (Figure 2.2C).

To determine if certain genic regions are associated with a specific methylation context, we examined the distribution of differentially hyper- and hypo-methylated genes in various annotated features of genic regions including promoter regions, 1000 bp upstream of the transcription start site (TSS), exons, and 5' and 3' untranslated regions (UTRs). The hyper- and hypo-methylation in CpG and CHG contexts occur predominantly in the gene body and to much less extent in the flanking regions, whereas CHH hyper-methylation was mainly located in the promoter regions (Figure 2.1D and E). A set of 45 genes showed both hyper- and hypo-methylation in various genic regions. To examine whether specific genomic regions are preferentially targeted for differential methylation during SCN infection, we mapped the differentially methylated regions to the 20 soybean chromosomes. The genomic distribution of the differentially methylated regions revealed that CpG and CHH regions were most enriched near the ends of the chromosomes (sub-telomeric regions) and to lesser extent in non-telomeric and centromeric regions (Figure 2.2A). In contrast, CHG regions were mainly localized in non-telomeric regions (Figure 2.2A), consistent with the preference of CHG methylation of targeting transposon-rich regions (Lister et al., 2008), which are located in the chromosome centers (Schmutz et al., 2010). The numbers of differentially hyper- and hypo-methylated regions in CpG, CHG and CHH were found to be significantly correlated with the chromosome size ($P < 0.05$). Thus, the differentially methylated regions seem to be equally distributed across the 20 chromosomes, taking into consideration chromosome size (Figure 2.2B and 2.2C).

It has been suggested that DNA methylation, particularly gene body methylation, exhibits a biased towards long genes versus short genes (Takuno and Gaut, 2013). Thus, we tested whether a correlation between DNA methylation/demethylation and gene size occurs. The methylation levels were plotted against the size of differentially hyper- and hypo-methylated genes. No bias towards long genes was observed in any sequences context, suggesting that methylation/demethylation occurs independently of gene size. Similarly, we asked whether GC content in the differentially methylated regions (200 bp) correlates with the methylation levels. The differential methylation levels in various sequence contexts were plotted against the GC content of the differentially hyper- and hypo-methylated regions. A constant profile of methylation levels versus various levels of GC content were observed, indicating that differential methylation occurs independent of GC content.

2.3 Recently duplicated soybean genes exhibit strong preference for differential methylation in response to SCN infection

It has been suggested that soybean genome experienced two whole genome-wide duplication (WGD) events approximately 59 and 13 million years ago (Mya) (Schmutz et al., 2010).We examined if differential methylation targets particularly genes contributed by a specific WGD event. We scanned soybean genome for duplicate blocks that contain the differentially hyper- and hypo-methylated genes and identified 673 collinearity events among the 703 hyper-methylated genes and 1,163 events among the 1,346 hypo-methylated genes. When we calculated the synonymous distance (*Ks*) values of these events, we found that recently duplicated genes contributed by the 13-Myr WGD (*Ks* = 0.06-0.39) are preferentially targeted for differential methylation compared to those genes contributed by the 59-Myr WGD (*Ks* = 0.40-0.80). Specifically, 2.71% (453 genes) of the paralogs contributed by the 13-Myr WGD were hyper-methylated ($\chi 2$ = 18.228, P = 3.946E-04) versus 2.2% (194 genes) of the paralogs contributed by the 59-Myr WGD ($\chi 2$ = 1.831, P = 0.608) (Figure 2.3A). Likewise, 4.68% (783 genes) of the

paralogs contributed by the 13-Myr WGD were hypo-methylated ($\chi 2$ = 32.194, P = 4.763E-07) versus 3.87% (341 genes) of the paralogs contributed by the 59-Myr WGD ($\chi 2$ = 2.025, P = 0.567) (Figure 2.3B). We further asked whether homologous genes are targeted similarly for differential methylation. Careful examination of the collinear genomic regions among the differentially hyper- and hypo-methylated genes resulted in the identification of 69 duplicated regions. These 69 duplicated regions contained 133 differentially methylated genes (Figure 2.3C). Homologs among these 133 genes are connected by red lines in Figure 2.3C.

2.4 Differential methylation impacts gene expression levels in SCN-infected roots

Recent studies have indicated that the effects of DNA methylation on gene expression levels differ depending on the genic regions and methylation context (Gent et al., 2013; Schmitz et al., 2013; Yang et al., 2015). Thus, we assayed the impact of DNA methylation in gene body and promoter regions on gene expression levels. RNA was isolated from the same SCN-infected and control samples and used in qPCR assays. A set of 24 genes (8 for each sequence context) was randomly selected to represent hyper- and hypo-methylation in gene body and promoter regions. Twenty-one of these 24 genes showed significant changes in mRNA expression levels in response to SCN infection relative to non-infected control (Figure 2.4 A-C). qPCR analysis also revealed an association between gene body hyper-methylation and both increased and decreased gene expression levels (Figure 2.4A). In contrast, hypo-methylation in gene body showed a general trend of increased gene expression levels (Figure 2.4C). Also, increased CpG, CHG and CHH methylation in the promoter regions showed positive or negative impacts on expression levels (Figure 2.4B), whereas demethylation showed a trend of increased gene expression levels (Figure 2.4D). These data suggest that differential methylation associated with SCN infection may have heterogeneous effects on gene expression levels.

2.5 Differentially methylated genes of SCN-infected roots are associated with various molecular functions

We classified the differentially methylated region-associated genes into different groups by molecular function using the Gene Ontology (GO) categorization from AgriGO database (Du et al., 2010) and conducted an enrichment analysis using Fisher's exact test (P <0.05). Molecular function groups corresponded to binding activity, transferase activity, catalytic activity, and hydrolase activity are significantly enriched among the differentially methylated genes. While it seems likely that methylation and demethylation impact various molecular function categories to a similar degree, we observed a significant difference in a number of molecular function groups between the differentially hyper- and hypo-methylated genes, specifically those associated with nucleic acid binding, RNA binding, RNA polymerase activity, hydrolase activity, microtubule and cytoskeletal protein binding, and kinase activity.

2.6 Identification of syncytium genes that are under methylation control

To identify syncytial genes that are under methylation control, we compared the differentially hyper- and hypo-methylated genes identified by us with our reference list of genes that change the expression in the syncytium induced by SCN (6903 genes) (Ithal et al., 2007; Klink et al., 2009; Klink et al., 2010; Kandoth et al., 2011). We found 70, 16 and 13 genes of the differentially hyper-methylated genes in CpG, CHG and CHH contexts, respectively, that overlapped with syncytial differentially expressed genes (Figure 2.5 A). Similarly, 123, 30 and 44 genes of the differentially hypo-methylated genes in CpG, CHG and CHH contexts, respectively were found to be overlapped with syncytial differentially expressed genes (Figure 2.5 B). After eliminating duplicated genes that were differentially methylated in more than one context, we identified 93 of the differentially hyper-methylated genes and 193 of the differentially hypo-methylated genes as overlapping with the 6,903 syncytium-regulated genes. Eight genes showed

both hyper-and hypo-methylation in different genic regions, resulting in 278 unique syncytium-differentially methylated genes. These 278 genes were classified using GO terms into various molecular function categories including binding activity, catalytic activity, transferase activity, kinase activity, and hydrolase activity (Figure 2.5 C). These categories were distributed to similar extent in hyper- and hypo-methylated genes overlapping with syncytium (Figure 2.5 D). Enrichment analysis revealed that genes involved in structural molecule activity and cofactor binding were significantly enriched among the 93 differentially hyper-methylated genes. Genes involved in sequence-specific DNA binding, hydrolase activity, and peptidase activity were significantly enriched among the 193 differentially hypo-methylated genes.

3. Discussion

We examined the impacts of SCN infection on the genome-wide DNA methylation patterns in soybean roots and identified both differentially methylated regions and the associated genes in CpG, CHH and CHG contexts. Our analysis provided several new insights into DNA methylation changes associated with SCN parasitism of soybean roots. We found that SCN induces hypo-methylation to much higher degree compared to hyper-methylation in all sequence contexts. This may explain the increased number of upregulated SCN-responsive genes relative to downregulated genes reported in various global gene expression studies (Ithal et al., 2007; Klink et al., 2009; Kandoth et al., 2011). A high proportion of demethylated cytosines has also been reported to occur in response to bacterial infection (Dowen et al., 2012). When the differentially methylated regions were assigned to overlapping protein-coding genes we found that methylation/demethylation occurs predominantly in the gene body in CpG and CHG contexts (Figure 2.1E and D), similar to the methylation patterns of Arabidopsis plants exposed to various biotic stresses (Dowen et al., 2012). The enrichment of non-CpG methylation in gen body is in contrast to recent findings in which gene body methylation was limited to CpG context (Cokus et al., 2008; Lister et al., 2008; Greaves et al., 2012; Takuno and Gaut, 2013). This controversy is presumably due to the fact that these studies used unchallenged plants. Whether differential gene body methylation in non-

CpG contexts is a specific signature of plants response to biotic stresses needs to be investigated further.

It has been demonstrated that variable methylation over the gene body is a general feature of plant and animal genomes (Cokus et al., 2008; Lister et al., 2008; Ball et al., 2009; Laurent et al., 2010). While the suite of functions of this methylation is not fully understood, it has been suggested that it might regulate alternative splicing efficiency and prevent aberrant transcription of long genes (Zilberman et al., 2007; Luco et al., 2010; Maunakea et al., 2010; Regulski et al., 2013). Gene body methylation/demethylation, specifically in CpG and CHH contexts, does not seem to severely impact gene expression in response to SCN infection. A moderate up- and down-regulation of differentially methylated genes was observed in our qPCR assays (Figure 2.4 A and B), suggesting that gene body methylation may fine-tune the transcriptional activity of targeted genes (Zilberman et al., 2007). The heterogeneous effect of body methylation on gene expression could be linked to the extent of methylation in the target genes. Moderate methylation tends to enhance gene expression, whereas low or high methylation tends to inhibit gene expression (Wang et al., 2013). The impact of gene body methylation on gene expression may also depend on the genic regions. Strong associations between DNA methylation in the first exon and gene downregulation have been reported in plants and animals (Brenet et al., 2011; Yang et al., 2015). First exon methylation was found to block transcript initiation and cause proximal polymerase pausing in human cells (Okitsu and Hsieh, 2007) and a similar mode of function has been suggested to occur in plants (Zilberman et al., 2007).

In addition to gene body methylation, we found the promoters of 447 genes were differentially methylated in CpG (153 genes), CHG (84 genes), and CHH (210 genes) contexts. Differential methylation particularly in the promoter regions may impact the accessibility of regulatory factors to their *cis* elements thereby regulating the transcriptional activity of SCN-responsive genes. Although DNA methylation in the promoter regions is often associated with decreased transcriptional activity, we observed a positive association between DNA methylation in the promoter and the expression levels of two genes in our qPCR analysis. Hyper- methylation in the

promoter region of these two genes may overlap with the binding sites of transcriptional repressor factors, where inefficient binding associated with increased DNA methylation may result in increased gene transcription. Similarly, CHH methylation in the promoter regions was found to positively impact gene expression in maize (Gent et al., 2013).

Interestingly, a small portion of the differentially methylated genes (126/2049) were found to be methylated in more than one sequence context simultaneously. This suggests that differential methylation in various contexts mostly occurs independently of each other, which is consistent with the fact that DNA methylation in CpG, CHG and CHH are mediated by three different enzymes including DNA METHYLTRANSFERASE 1 (MET1), the plant-specific CHROMOMETHYLASE 3 (CMT3), and DOMAINS REARRANGED METHYLTRANSFERASEs (DRMs), respectively (Lindroth et al., 2001; Cao and Jacobsen, 2002; Kankel et al., 2003). Although the biological significance of simultaneous methylation in various sequence contexts remains an important question, a coordinated function of these enzymes may occur and could involve other epigenetic regulating factors and various forms of RNA polymerase activity.

SCN dramatically alters the transcription machinery of infected host cells (Klink et al., 2009; 2010; Kandoth et al., 2011), a process that may generate aberrant transcripts. These aberrant transcripts are processed by the RNA-induced silencing complex (RISC) into siRNAs, which can induce DNA methylation through the RdDM pathway (Chan et al., 2005; Henderson and Jacobsen, 2007). The finding that 24-nt small RNA reads are the most abundant class among other small RNA populations in nematode infected root samples (Hewezi et al., 2008, Li et al., 2012) supports this hypothesis. SCN infection seems to trigger systemic hyper- and hypo methylation in root cells that are not in direct contact with the nematodes. This is consistent with our finding that several of the differentially methylated genes, identified using the whole root system, change the expression in response to SCN infection but only 4% (278/6,903) of these genes were among the syncytium differentially expressed genes. The mechanism through which SCN induces methylation changes in a systemic manner is unknown but this process may involve active gene expression changes of main epigenetic regulating genes and SCN-responsive siRNAs. Several epigenetic regulating genes were identified as

differentially expressed in the SCN syncytium (Ithal et al., 2007; Klink et al., 2010; Kandoth et al., 2011). In addition, SCN induces differential accumulation of siRNAs (Li et al., 2012), and hence siRNA may function in a non-cell-autonomous fashion similar to the mobile epigenetically activated siRNAs in pollen (Slotkin et al., 2009) to shape the epigenome of non-syncytial cells.

One striking finding of our analysis is that differential methylation was significantly enriched in the recently duplicated genes compared to early duplicated genes. In agreement with this finding, it has been recently reported that differential methylation in DNA occur in recently duplicated region of the genome (Schmitz et al., 2013; Keller and Soojin, 2014). This pattern of methylation differences may contribute to gene dosage rebalance, thereby providing cell-type specific regulation. Targeting recently duplicated genes for differential methylation by SCN could be due to the plausible association of the recently duplicated genes with disease and stress responses over the ancient copies, which are assumed to be involved in more basic and less stress-related cellular processes (Kondrashov, 2012).

Our finding that only 4% of the syncytium differentially expressed genes are differentially methylated suggests that regulation of gene expression in syncytium is heavily influenced by the well-known transcription factor-based regulatory mechanisms. Remarkably, we found several genes involved in DNA methylation and epigenetic regulation among the syncytium differentially methylated genes, including homologs of Arabidopsis O-acetyltransferases, S-adenosyl-L-methionine-dependent methyltransferases protein, dicer-like 4, DEFECTIVE IN MERISTEM SILENCING 2 (DMS2), SU(VAR)3-9 homolog 9, chromatin remodeling complex subunit B, chromatin remodeling factor17, and a PHD finger family protein. This indicates that methylation changes in the syncytium are tightly controlled and a feedback loop may exist between various epigenetic components that regulate gene expression changes in the syncytium. A feedback regulatory loop between various epigenetic components has been recently suggested (Du et al., 2012).

Differential methylation induced by SCN in the syncytium seems to impact genes associated with various molecular functions. Suppression of defense responses is one of the main characteristics of the compatible interaction between SCN and soybean (Ithal et al., 2007; Hosseini and Matthews, 2014). Though a high proportion of defense-related genes was found to be down-regulated in the syncytium, only a few of these genes were differentially methylated. This is in contrast to the finding that differential methylation induced by the bacterial pathogen *Pst* in Arabidopsis is preferentially associated with genes involved in defense responses (Dowen et al., 2012; Yu et al., 2013).

Differential methylation may also contribute to the regulation of phytohormone-mediated soybean susceptibility to SCN. Various genes involved in auxin signaling were among the differentially methylated genes in the syncytium, including an auxin receptor homolog, auxin response factor 8, and indole-3-acetic acid inducible 9 (IAA9). Auxin plays crucial roles in the initiation and formation of the syncytium (Grunewald et al., 2009; Gheysen and Mitchum, 2011; Hewezi et al., 2014), and regulation of auxin signaling by DNA methylation was recently reported in Arabidopsis (Li et al., 2011). In addition, a homolog to Arabidopsis histone deacetylase1 (HDA19), which is involved in jasmonic acid and ethylene dependent pathogen resistance was differentially methylated (Zhou et al., 2005; Choi et al., 2012). Another homolog to Arabidopsis ethylene insensitive 3 (EIN3) family protein was also among the syncytium hypo-methylated genes. EIN3 is a transcription factor that mediates downstream transcriptional cascades for ethylene responses (Chao et al., 1997; Binder et al., 2007) and various Arabidopsis EIN mutants showed reduced susceptibility to the cyst nematode *H. schachtii* (Wubben et al., 2001; Wubben et al., 2004).

The plant cytoskeleton plays fundamental role in mediating changes in the morphology and structure of the syncytium (de Almeida Engler and Favery, 2011) and DNA demethylation may contribute to the regulation of this process. Genes involved in the dynamics of cytoskeleton, including beta tubulin, microtubule-associated protein and NIMA-related serine/threonine kinase were identified as differentially methylated in the syncytium. Similarly, cell wall biogenesis is extensively altered during syncytium

initiation and expansion (Golinowski et al., 1996; Wieczorek et al., 2008; Siddique et al., 2009; Tucker et al., 2011). Several genes coding for enzymes with functions related to cell wall architecture and remodeling showed differential methylation accompanied with gene expression changes in the syncytium. This included, for example, cellulose synthase, pectate lyases, galacturonosyltransferase, and beta galactosidase.

Signal transduction and regulatory genes, such as those encoding transcription factors and protein kinases, represent a significant portion of syncytium differentially methylated genes. Transcription factors belonging to C2H2, bZIP, MYB, WRKY, F-box/RNI-like, calmodulin-binding transcription activator, B-Box Zinc Finger, CPP, and HB-PHD family proteins were differentially methylated in the syncytium. Members of some of these gene families have been shown to be involved in defense signaling pathways (Eulgem and Somssich, 2007; Reddy et al., 2011; Ambawat et al., 2013). In addition, a high number of protein kinases were highly represented among the syncytium differentially methylated genes, including components of MAP kinase signaling cascades. Thus, DNA methylation may be of particular importance in defining signal specificity associated with SCN parasitism. An important reprogramming of plant primary metabolism also occurs throughout syncytium formation and nematode feeding (Hofmann et al., 2010). Several genes involved in primary metabolism pathways, specifically those related to carbohydrates degradation and glycolysis, were differentially methylated. The potential transcriptional regulation of metabolic processes by DNA methylation may indicate that metabolite levels in the syncytium are tightly controlled.

Recent experimental evidence indicated that various plant pathogens have evolved the ability to manipulate host ubiquitin/proteasome system (UPS) to cause disease (DIELEN et al., 2010). UPS seems to play a vital role in syncytium formation as many genes with UPS-related functions were found to be significantly up-regulated in the developing syncytia during a compatible interaction between *H. glycines* and soybean (Ithal et al., 2007; Klink et al., 2007). Demethylation of many of these genes, including for example proteasome component (PCI) domain protein, MATH-BTB domain protein (BPM), and ubiquitin protein ligase, may be the cause of their up-regulation in the syncytium. The finding that cyst nematodes secrete effector proteins that may function

in dysregulation of host UPS (Tytgat et al., 2004; Chronis et al., 2013), may point to a role of differential methylation as an additional mechanism of control to regulate particular cellular processes targeted by nematode effectors to promote parasitism.

In conclusion, our genome-wide DNA methylation approach provide unprecedented insights into DNA methylation changes during the compatible interaction of SCN with soybean that impact a large number of protein coding genes, locally in the syncytium and systemically in non-infected root cells, and their eventual transcriptional regulation.

4. Materials and Methods

4.1 Plant and nematode material

Soybean (Glycine max (L.) Merr) cv. Williams 82, an SCN-susceptible cultivar, was used in this study. SCN (*H. glycines* Ichinohe) race 3 (HG-type 0) was maintained in greenhouse culture on Williams 82 using standard procedures (Niblack et al., 2009).

4.2 Nematode inoculation and tissue collection

Soybean seed were soaked in running tap water for 30 min, surface-sterilized in 10% bleach solution for 10 min, followed by thorough rinsing in running tap water. Seed were germinated for 3 days in rolled germination paper (rag dolls) at 26°C in the dark. Freshly hatched second-stage juveniles (J2) were extensively washed and suspended in 0.1% sterile agarose at a concentration of 400 J2 per 100 µl. Healthy 3-d-old seedlings were then placed on moist blue blotter paper in the petri dish (150 mm ×15 mm) and inoculated with about 2000 J2 distributed equally around the whole roots. Control samples were mock inoculated with 500 µl of 0.1% agarose. SCN-inoculated and control roots were covered with sterile blue blotter paper soaked with 2-(N-

morpholino)ethanesulfonic acid (MES)-buffered sterile water (pH 6.5). The Petri dishes were then covered with the lids and incubated in a plant growth chamber at 26°C in 16-h-light (75µmol m^{-2} s^{-1})/8-h-dark conditions. The infection process was examined in 20% of the inoculated samples using acid fuschin stain as previously described by (Daykin and Hussey, 1985). The inoculation experiment was repeated three times and two root samples (infected and control) from each experiment for a total of six biologically independent samples that were collected at five days post infection (dpi). Each sample contained at least 6 seedlings.

4.3 Construction of Methyl-Seq Libraries

Whole genome Methyl-Seq Libraries were prepared using the Illumina TruSeq Library Prep kit (Illumina, San Diego, CA) with minor modifications. Briefly, DNA was extracted from root tissues of control and inoculated seedling using DNAeasy Plant Mini Kit (Qiagen). Approximately, 1.5 µg of genomic DNA plus unmethylated lambda DNA were sheared using the Bioruptor (Diagenode Inc) as per manufacturer instructions. Sheared genomic DNA was spiked with 1-2% of fragmented, unmethylated lambda DNA (Promega). Fragment size distribution of 250-400 bp for each sample was verified by a run on Agilent Bioanalyzer using a 1000 DNA chip (Agilent Technologies). Cytosine methylated adapters provided by Illumina were ligated to the blunt ends of fragmented DNA. The ligated DNA was treated with sodium bisulfite using MethylCode™ Bisulfite Conversion Kit (Invitrogen) and ligated fragments (400-500 bp) were selected on a Pippin prep (Sage Sciences). Post-size selection, samples were enriched by 10 cycles of PCR reaction using 2.5U uracil-insensitive PfuTurboC_xHotstart polymerase (Agilent), 5 µl 10X PfuTurbo Buffer, 0.4 µl of 100 nM dNTPs, 5 µl of Illumina Truseq oligo mix. The PCR reactions were run using the following program: 95°C for 2 min, and 10 cycles of 98°C for 15 sec, 60°C for 30 sec, 72°C for 4 min, followed by a final elongation step of 10 min at 72°C. The amplification products were cleaned using Agencourt AMPure XP beads (Beckman Coulter, Inc). Purified PCR products were enriched further for 5 cycles of PCR amplification using Illumina's protocol for library enrichment. Finally, Methyl-

Seq Libraries were validated on Agilent Bioanalyzer using a 1000 DNA chip (Agilent Technologies) and quantified using Illumina sequencing primers in a qPCR reaction before sequencing. Sequencing was performed at Oak Ridge National Laboratory using the Illumina HiSeq platform.

4.4 Data Analysis

MethylC-Seq reads for each biological replicate were filtered and trimmed using Trimmomatic with a phred quality threshold of 33 (Bolger et al., 2014). All sequencing reads were individually mapped using Bismark (Krueger and Andrews, 2011) to the soybean genome reference (Glycine max Wm82.a2.v1). The alignments were done using the default software options for paired end read alignment. The SAM alignment files generated by Bismark were used to identify differentially methylated cytosine bases or regions using the methylKit package in R with logistic regression (Akalin et al., 2012). Cytosines were called if the genomic region was covered by at least 10 reads and three methylation call files for CpG, CHG and CHH contexts were generated for both infected and control samples. Differentially methylated regions were identified using a non-overlapping window of 200 bp. Hypo/hyper-methylated regions covered by at least 10 reads with percent methylation difference larger than 25%, were identified using q-value <0.01. Fisher's exact test was used to calculate P-values and P-values were adjusted to q-values using SLIM method (Wang et al., 2011). The location and function of genes in the soybean genome version Wm82.a2.v1 were downloaded from Phytozome (Goodstein et al., 2012). Differentially methylated regions were mapped to various annotated features of genic regions including protein coding genes their subfeatures, including promoter regions, exons, and 5' and 3' untranslated regions (UTRs), using a custom R script and the Bioconductor package rtracklayers (Lawrence et al., 2009). The promoter region was defined as 1000 bases upstream of the first base of the mRNA annotated region. A positive overlap was recorded if the 200 bp region overlapped the feature or subfeature by one base pair or more. Gene Ontology (GO) enrichment analysis was performed using AgriGO database (Du et al., 2010) and statistical significance was determined

using Fisher's exact test and a *P*-value cutoff less than 0.05. Enrichment analyses of the differentially methylated genes were conducted using the latest annotated version of soybean genome (*Glycine max* Wm82.a2.v1) as background genes. Enrichment analyses of the differentially hyper- and hypo-methylated genes that overlap with the syncytium differentially expressed genes were conducted using the total number of differentially methylated genes (2004 genes) as background genes. χ^2 tests of independence were used to evaluate the significance of overlaps between various gene lists as previously described by Hewezi et al. (2012).

4.5 RNA isolation and qPCR quantifications

Total RNA was extracted from frozen ground root tissues using the method previously described by (Verwoerd et al., 1989). Total RNA was treated with DNase using DNase I kit (New England Biolabs, Ipswitch, Massachusetts, USA). 50 ng of DNase-treated RNA were used for cDNA synthesis and PCR amplification using Verso one- step qPCR SYBR kit (Fisher Scientific, Hampton, New Hampshire, USA) according to the manufacturer's protocol. The reactions were run on ABI 7900HT Fast Real-Time PCR System (Applied Biosystems) using the following program: 50°C for 15 min, 95°C for 15 min, and 40 cycles of 95°C for 15 s, 60°C for 30 s and 72°C for 20 s. A dissociation curve was created after PCR amplification to detect nonspecific products using the following program: 95°C for 15 s, 50°C for 15 s and a slow ramp from 50°C to 95°C. Three biologically independent samples of SCN-infected and non-infected control each with four technical replicates were used in qPCR analysis. The soybean ubiquitin (Glyma.20G141600), a constitutively expressed gene, was used as an internal control to normalize gene expression levels. Quantification of gene expression changes in SCN-infected samples relative to non-infected control were performed as previously described by (Hewezi et al., 2015).

References

Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, Mason CE (2012) methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biology **13: R87**

Akimoto K, Katakami H, Kim H-J, Ogawa E, Sano CM, Wada Y, Sano H (2007) Epigenetic inheritance in rice plants. Annals of botany **100:** 205-217

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

Alvarez ME, Nota F, Cambiagno DA (2010) Epigenetic control of plant immunity. Molecular Plant Pathology **11:** 563-576

Ambawat S, Sharma P, Yadav NR, Yadav RC (2013) MYB transcription factor genes as regulators for plant responses: an overview. Physiology and Molecular Biology of Plants **19:** 307-321

Ball MP, Li JB, Gao Y, Lee J-H, LeProust EM, Park I-H, Xie B, Daley GQ, Church GM (2009) Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. Nature biotechnology **27:** 361-368

Becker C, Hagmann J, Müller J, Koenig D, Stegle O, Borgwardt K, Weigel D (2011) Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. Nature **480:** 245-249

Binder BM, Walker JM, Gagne JM, Emborg TJ, Hemmann G, Bleecker AB, Vierstra RD (2007) The Arabidopsis EIN3 binding F-Box proteins EBF1 and EBF2 have distinct but overlapping roles in ethylene signaling. The Plant **19:** 509-523

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30:** 2114–2120

Boyko A, Kathiria P, Zemp FJ, Yao Y, Pogribny I, Kovalchuk I (2007) Transgenerational changes in the genome stability and methylation in pathogen-

infected plants (virus-induced plant genome instability). Nucleic Acids Research **35:** 1714-1725

Brenet F, Moh M, Funk P, Feierstein E, Viale AJ, Socci ND, Scandura JM (2011) DNA methylation of the first exon is tightly linked to transcriptional silencing. PLoS One **6:** e14524.

Cao X, Aufsatz W, Zilberman D, Mette MF, Huang MS, Matzke M, Jacobsen SE (2003) Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. Current Biology **13:** 2212-2217

Cao X, Jacobsen SE (2002) Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. Proceedings of the National Academy of Sciences **99:** 16491-16498

Chan SW, Henderson IR, Jacobsen SE (2005) Gardening the genome: DNA methylation in *Arabidopsis thaliana*. Nature Reviews Genetics **6:** 351–360

Chao Q, Rothenberg M, Solano R, Roman G, Terzaghi W, Ecker JR (1997) Activation of the ethylene gas response pathway in Arabidopsis by the nuclear protein ETHYLENE-INSENSITIVE3 and related proteins. Cell **89:** 1133-1144

Choi SM, Song HR, Han SK, Han M, Kim CY, Park J, Lee YH, Jeon JS, Noh YS, Noh B (2012) HDA19 is required for the repression of salicylic acid biosynthesis and salicylic acid-mediated defense responses in Arabidopsis. The Plant Journal **71:** 135-146

Chronis D, Chen S, Lu S, Hewezi T, Carpenter SC, Loria R, Baum TJ, Wang X (2013) A ubiquitin carboxyl extension protein secreted from a plant-parasitic nematode *Globodera rostochiensis* is cleaved in planta to promote plant parasitism. The Plant Journal **74:** 185-196

Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature **452:** 215-219

Cook DE, Bayless AM, Wang K, Guo X, Song Q, Jiang J, Bent AF (2014) Distinct copy number, coding sequence, and locus methylation patterns underlie

Rhg1-mediated soybean resistance to soybean cyst nematode. Plant physiology **165:** 630-647

Daykin M, Hussey R (1985) Staining and histopathological techniques in nematology. In: Barker KR, Carter CC, Sasser JN, editors. An Advanced treatise on Meloidogyne: Volume II. Raleigh, NC: North Carolina State University Graphics; 1985. pp. 39–48.

de Almeida Engler J, Favery B (2011) The plant cytoskeleton remodelling in nematode induced feeding sites. *In* Genomics and Molecular Genetics of Plant-Nematode Interactions. Springer, pp 369-393

Dielen AS, Badaoui S, Candresse T, German-Retana S (2010) The ubiquitin/26S proteasome system in plant–pathogen interactions: A never-ending hide-and-seek game. Molecular Plant Pathology **11:** 293-308

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

Du J, Zhong X, Bernatavichute YV, Stroud H, Feng S, Caro E, Vashisht AA, Terragni J, Chin HG, Tu A (2012) Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. Cell **151:** 167-180

Du Z, Zhou X, Ling Y, Zhang Z, Su Z (2010) agriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Res **38:** W64–70.

Eulgem T, Somssich IE (2007) Networks of WRKY transcription factors in defense signaling. Current Opinion in Plant Biology **10:** 366-371

Gent JI, Ellis NA, Guo L, Harkess AE, Yao Y, Zhang X, Dawe RK (2013) CHH islands: de novo DNA methylation in near-gene chromatin regulation in maize. Genome Research **23:** 628-637

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

Golinowski W, Grundler F, Sobczak M (1996) Changes in the structure ofArabidopsis thaliana during female development of the plant-parasitic nematode *Heterodera schachtii*. Protoplasma **194:** 103-116

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research **40:** D1178-D1186

Greaves IK, Groszmann M, Ying H, Taylor JM, Peacock WJ, Dennis ES (2012) Trans chromosomal methylation in Arabidopsis hybrids. Proceedings of the National Academy of Sciences **109:** 3570-3575

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

He G, Elling AA, Deng XW (2011) The epigenome and plant development. Annual Review of Plant Biology **62:** 411-435

Henderson IR, Jacobsen SE (2007) Epigenetic inheritance in plants. Nature **447:** 418-424

Hewezi T, Baum TJ (2013) Manipulation of plant cells by cyst and root-knot nematode effectors. Molecular Plant-Microbe Interactions **26:** 9-16

Hewezi T, Howe P, Maier TR, Baum TJ (2008) Arabidopsis small RNAs and their targets during cyst nematode parasitism. Molecular Plant-Microbe Interactions **21:** 1622-1634

Hewezi T, Juvale PS, Piya S, Maier TR, Rambani A, Rice JH, Mitchum MG, Davis EL, Hussey RS, Baum TJ (2015) The Cyst Nematode Effector Protein 10A07 Targets and Recruits Host Posttranslational Machinery to Mediate Its Nuclear Trafficking and to Promote Parasitism in Arabidopsis. The Plant Cell **27:** 891-907

Hewezi T, Piya S, Richard G, Rice JH (2014) Spatial and temporal expression patterns of auxin response transcription factors in the syncytium induced by the beet cyst nematode *Heterodera schachtii* in Arabidopsis. Molecular Plant Pathology **15:** 730-736

Hofmann J, El Ashry AEN, Anwar S, Erban A, Kopka J, Grundler F (2010) Metabolic profiling reveals local and systemic responses of host plants to nematode parasitism. The Plant Journal **62:** 1058-1071

Hosseini P, Matthews BF (2014) Regulatory interplay between soybean root and soybean cyst nematode during a resistant and susceptible reaction. BMC Plant Biology **14:** 300

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

Kandoth PK, Ithal N, Recknor J, Maier T, Nettleton D, Baum TJ, Mitchum MG (2011) The soybean Rhg1 locus for resistance to the soybean cyst nematode Heterodera glycines regulates expression of a large number of stress-and defense-related genes in degenerating feeding cells. Plant Physiology **155:** 1960-1975

Kankel MW, Ramsey DE, Stokes TL, Flowers SK, Haag JR, Jeddeloh JA, Riddle NC, Verbsky ML, Richards EJ (2003) Arabidopsis MET1 cytosine methyltransferase mutants. Genetics **163:** 1109-1122

Keller TE, Soojin VY (2014) DNA methylation and evolution of duplicate genes. Proceedings of the National Academy of Sciences **111:** 5932-5937

Khan R, Alkharouf N, Beard H, MacDonald M, Chouikha I, Meyer S, Grefenstette J, Knap H, Matthews B (2004) Microarray analysis of gene expression in soybean roots susceptible to the soybean cyst nematode two days post invasion. Journal of Nematology **36:** 241-248

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

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

Klink VP, Overall CC, Alkharouf NW, MacDonald MH, Matthews BF (2007) Laser capture microdissection (LCM) and comparative microarray expression analysis of syncytial cells isolated from incompatible and compatible soybean

(*Glycine max*) roots infected by the soybean cyst nematode (Heterodera glycines). Planta **226:** 1389-1409

Koenning SR, Wrather JA (2010) Suppression of soybean yield potential in the continental United States by plant diseases from 2006 to 2009. Plant Health Program, doi:10.1094/PHP-2010-122-01-RS.

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

Krueger F, Andrews SR (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics **27:** 1571-1572

Laurent L, Wong E, Li G, Huynh T, Tsirigos A, Ong CT, Low HM, Sung KWK, Rigoutsos I, Loring J (2010) Dynamic changes in the human methylome during differentiation. Genome Research **20:** 320-331

Lawrence M, Gentleman R, Carey V (2009) rtracklayer: an R package for interfacing with genome browsers. Bioinformatics **25:** 1841-1842

Li W, Liu H, Cheng ZJ, Su YH, Han HN, Zhang Y, Zhang XS (2011) DNA methylation and histone modifications regulate de novo shoot regeneration in Arabidopsis by modulating WUSCHEL expression and auxin signaling. PLoS Genetics **7:** e1002243

Li X, Wang X, Zhang S, Liu D, Duan Y, Dong W (2012) Identification of soybean microRNAs involved in soybean cyst nematode infection by deep sequencing. PLoS One **7:** e39650

Lindroth AM, Cao XF, Jackson JP, Zilberman D, McCallum CM, Henikoff S, Jacobsen SE (2001) Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. Science **292:** 2077-2080

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell **133:** 523-536

Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T (2010) Regulation of alternative splicing by histone modifications. Science **327:** 996-1000

Matzke M, Kanno T, Daxinger L, Huettel B, Matzke AJ (2009) RNA-mediated chromatin-based silencing in plants. Current Opinion in Cell Biology **21:** 367-376

Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, Johnson BE, Hong C, Nielsen C, Zhao Y (2010) Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature **466:** 253-257

Niblack T, Tylka GL, Arelli P, Bond J, Diers B, Donald P, Faghihi J, Ferris VR, Gallo K, Heinz RD, Lopez-Nicora H, Von Qualen R, Welacky T, Wilcox J (2009) A standard greenhouse method for assessing soybean cyst nematode resistance in soybean: sCE08 (standardized cyst evaluation 2008). Plant Health Progress. doi:10.1094/PHP-2009-0513-01-RV.

Okitsu CY, Hsieh CL (2007) DNA methylation dictates histone H3K4 methylation. Mol Cell Biol **27:** 2746–2757.

Reddy AS, Ali GS, Celesnik H, Day IS (2011) Coping with stresses: roles of calcium- and calcium/calmodulin-regulated gene expression. The Plant Cell **23:** 2010-2032

Regulski M, Lu Z, Kendall J, Donoghue MT, Reinders J, Llaca V, Deschamps S, Smith A, Levy D, McCombie WR (2013) The maize methylome influences mRNA splice sites and reveals widespread paramutation-like switches guided by small RNA. Genome Research **23:** 1651-1662

Schmitz RJ, He Y, Valdés-López O, Khan SM, Joshi T, Urich MA, Nery JR, Diers B, Xu D, Stacey G (2013) Epigenome-wide inheritance of cytosine methylation variants in a recombinant inbred population. Genome Research **23:** 1663-1674

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J (2010) Genome sequence of the palaeopolyploid soybean. Nature **463:** 178-183

Siddique S, Endres S, Atkins JM, Szakasits D, Wieczorek K, Hofmann J, Blaukopf C, Urwin PE, Tenhaken R, Grundler FM (2009) Myo-inositol oxygenase genes are involved in the development of syncytia induced by *Heterodera schachtii* in Arabidopsis roots. New Phytologist **184:** 457-472

Slotkin RK, Vaughn M, Borges F, Tanurdžić M, Becker JD, Feijó JA, Martienssen RA (2009) Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell **136:** 461-472

Takuno S, Gaut BS (2013) Gene body methylation is conserved between plant orthologs and is of evolutionary consequence. Proceedings of the National Academy of Sciences **110:** 1797-1802

Tucker ML, Murphy CA, Yang R (2011) Gene expression profiling and shared promoter motif for cell wall modifying proteins expressed in soybean cyst nematode infected roots. Plant Physiology**156:** 319–329

Tytgat T, Vanholme B, De Meutter J, Claeys M, Couvreur M, Vanhoutte I, Gheysen G, Van Criekinge W, Borgonie G, Coomans A (2004) A new class of ubiquitin extension proteins secreted by the dorsal pharyngeal gland in plant parasitic cyst nematodes. Molecular Plant-Microbe Interactions **17:** 846-852

Vaughn MW, Tanurdžić M, Lippman Z, Jiang H, Carrasquillo R, Rabinowicz PD, Dedhia N, McCombie WR, Agier N, Bulski A (2007) Epigenetic natural variation in Arabidopsis thaliana. PLoS Biology **5:** e174

Verwoerd TC, Dekker BMM, Hoekema A (1989) A Small-Scale Procedure for the Rapid Isolation of Plant Rnas. Nucleic Acids Research **17:** 2362-2362

Wang H-Q, Tuominen LK, Tsai C-J (2011) SLIM: a sliding linear model for estimating the proportion of true null hypotheses in datasets with dependence structures. Bioinformatics **27:** 225-231

Wang Y, Wang X, Lee TH, Mansoor S, Paterson AH (2013) Gene body methylation shows distinct patterns associated with different gene origins and duplication modes and has a heterogeneous relationship with gene expression in Oryza sativa (rice). New Phytologist **198:** 274-283

Wieczorek K, Hofmann J, Blöchl A, Szakasits D, Bohlmann H, Grundler FM (2008) Arabidopsis endo-1,4-β-glucanases are involved in the formation of root syncytia induced by Heterodera schachtii. The Plant Journal **53:** 336-351

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

Wubben MJE, Rodermel SR, Baum TJ (2004) Mutation of a UDP-glucose-4-epimerase alters nematode susceptibility and ethylene responses in Arabidopsis roots. Plant Journal **40:** 712-724

Wubben MJE, Su H, Rodermel SR, Baum TJ (2001) Susceptibility to the sugar beet cyst nematode is modulated by ethylene signal transduction in *Arabidopsis thaliana*. Molecular Plant-Microbe Interactions **14:** 1206-1212

Yang HX, Chang F, You CJ, Cui J, Zhu GF, Wang L, Zheng Y, Qi J, Ma H (2015) Whole-genome DNA methylation patterns and complex associations with gene structure and expression during flower development in Arabidopsis. Plant Journal **81:** 268-281

Yang SM, Tang F, Caixeta ET, Zhu HY (2013) Epigenetic regulation of a powdery mildew resistance gene in *Medicago truncatula*. Molecular Plant **6:** 2000-2003

Yu A, Lepere G, Jay F, Wang JY, Bapaume L, Wang Y, Abraham AL, Penterman J, Fischer RL, Voinnet O, Navarro L (2013) Dynamics and biological relevance of DNA demethylation in Arabidopsis antibacterial defense. Proceedings of the National Academy of Sciences of the United States of America **110:** 2389-2394

Zhang MS, Kimatu JN, Xu KZ, Liu B (2010) DNA cytosine methylation in plant development. Journal of Genetics and Genomics **37:** 1-12

Zhang XY, Yazaki J, Sundaresan A, Cokus S, Chan SWL, Chen HM, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, Ecker JR (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in Arabidopsis. Cell **126:** 1189-1201

Zhou CH, Zhang L, Duan J, Miki B, Wu KQ (2005) HISTONE DEACETYLASE19 is involved in jasmonic acid and ethylene signaling of pathogen response in Arabidopsis. The Plant Cell **17:** 1196-1204

Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S (2007) Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. Nature Genetics **39:** 61-69

Appendix

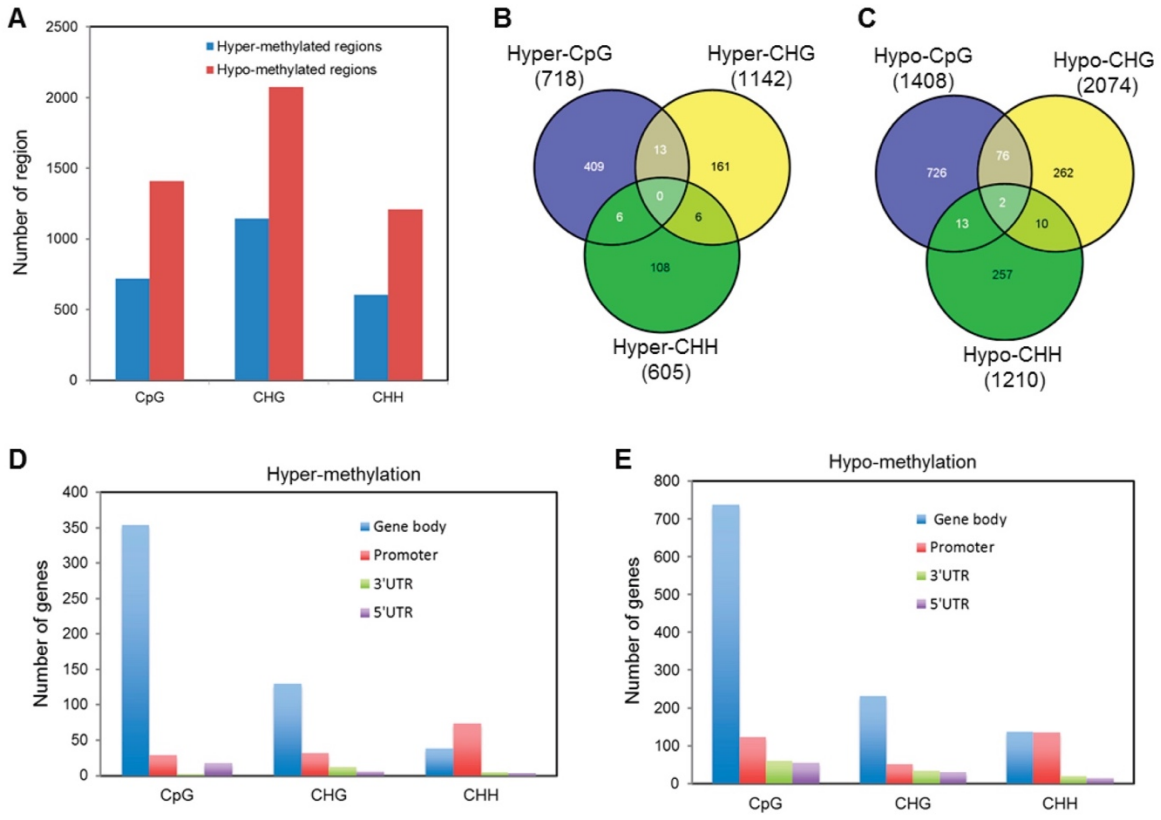

Figure 2.1 Characterization of differentially methylated regions in soybean roots in response to SCN infection.

(A) Number of differentially hyper- and hypo-methylated regions between SCN-infected and non-infected roots in CpG, CHG and CHH contexts. (B and C) Number of protein-coding genes overlapping with differentially hyper-methylated regions (B) or with hypo-methylated regions (C) in CpG, CHG and CHH contexts. The total number of differentially methylated regions in each context is shown in parentheses. (D and E) Distribution of differentially hyper-methylated (D) and Hypo-methylated (E) regions to various annotated features of protein–coding genes.

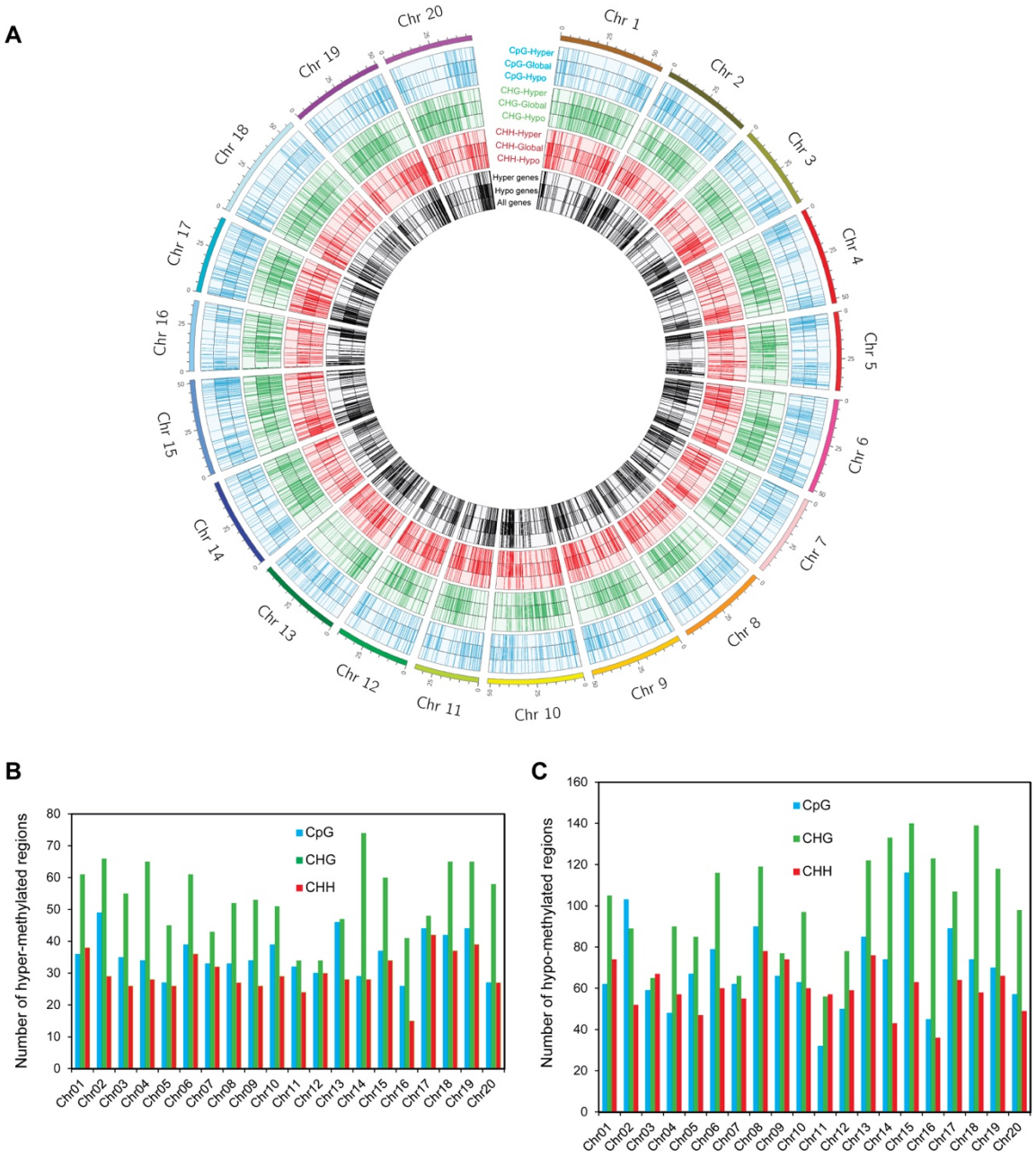

Figure 2.2 Genome-wide mapping of differentially methylated regions and associated genes. **(A)** A circle diagram showing chromosomal locations of the differentially hyper- and hypo-methylated regions in CpG, CHG and CHH contexts and overlapping genes into the 20 soybean chromosomes. (B and C) Distribution of the differentially hyper-methylated regions (B) and hypo-methylated regions (C) in the 20 soybean chromosomes.

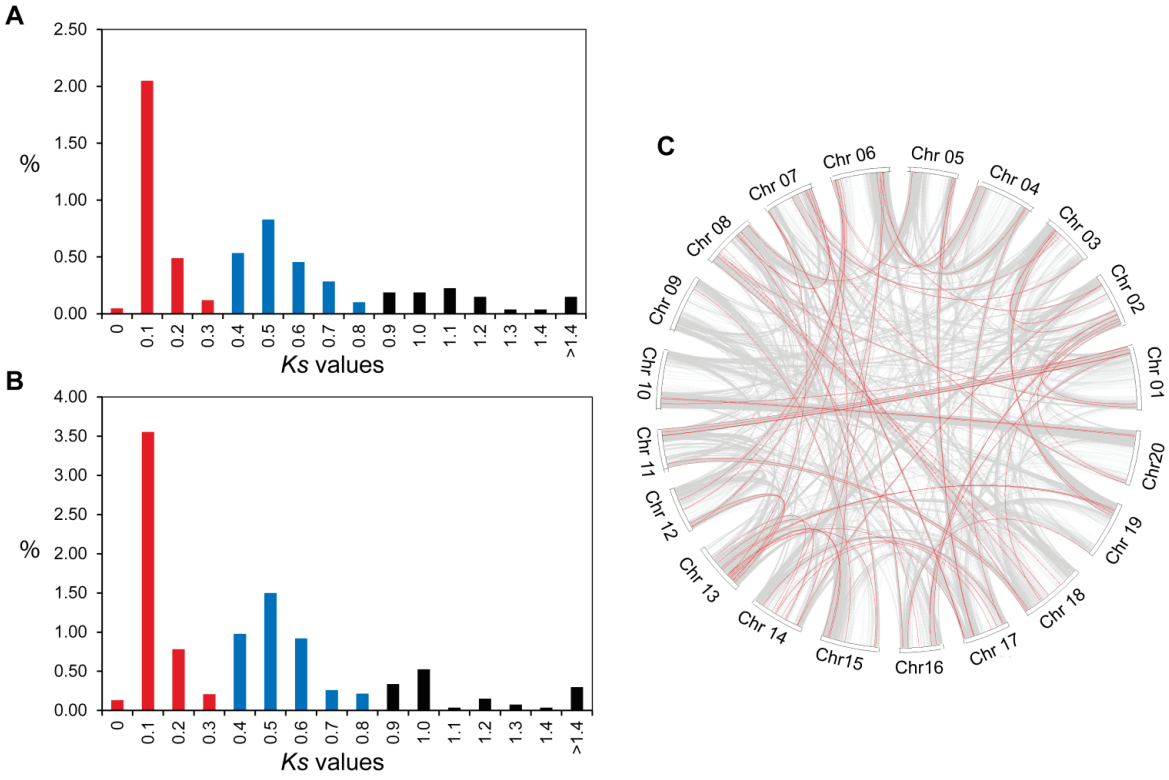

Figure 2.3 Recently duplicated soybean genes are preferentially targeted for differential methylation in response to SCN infection. (A) Percentages of hyper-methylated genes with various Ks values relative to the total number of paralogs that were contributed by the 13-Myr WGD (red bars; Ks = 0.06-0.39) or by the 59-Myr WGD (blue bars; Ks = 0.40-0.80). (B) Percentages of the hypo-methylated genes with various Ks values relative to the total number of paralogs that were contributed by the 13-Myr or 59-Myr WGD events. The Ks values for all differentially methylated genes were calculated using MCScanX program. (C) Homologous genes were targeted correspondingly for differential methylation. Shown are 69 collinearity events (red lines) among 133 differentially methylated genes. Grey lines represent whole genome duplicated genes contributed by the 13- and 59-Myr WGD events.

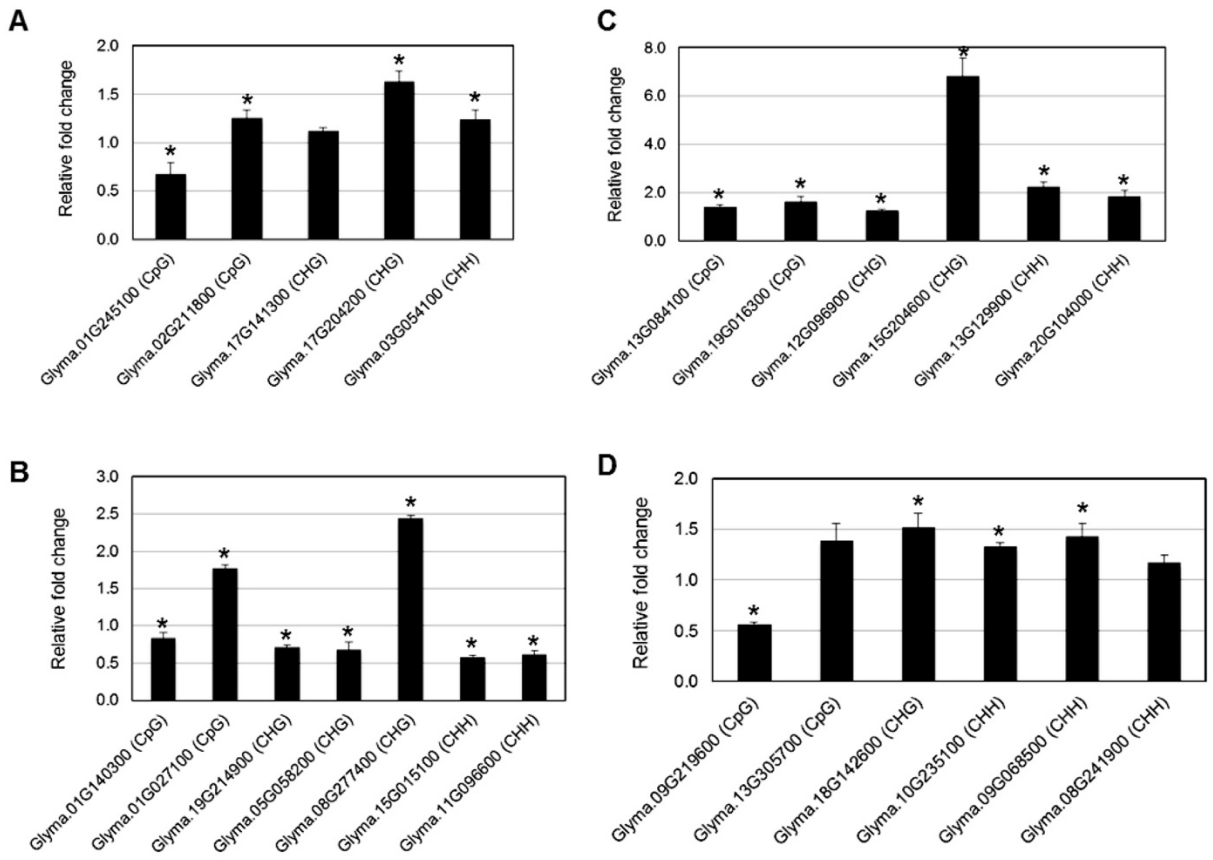

Figure 2.4 qPCR quantification of the impact of DNA methylation in gene body and promoter regions on gene expression levels. (A and B) Hyper-methylation in gene body (A) or promotor regions (B) in various sequence contexts showed heterogeneous effects on gene expression levels. (C and D) Hypo-methylation in gene body (C) or promotor regions (D) in various sequence contexts was mainly associated with increased gene expression levels. The relative fold-change values were calculated using the 2−ΔΔCT method and represent changes of the expression levels in SCN-infected root tissues relative to non-infected controls. Data are averages of three biologically independent experiments ± SE. Mean values significantly different from 1.0 (no change) are indicated by an asterisk as determined by t tests ($P < 0.05$). Soybean ubiquitin (Glyma.20G141600) was used as an internal control to normalize gene expression levels.

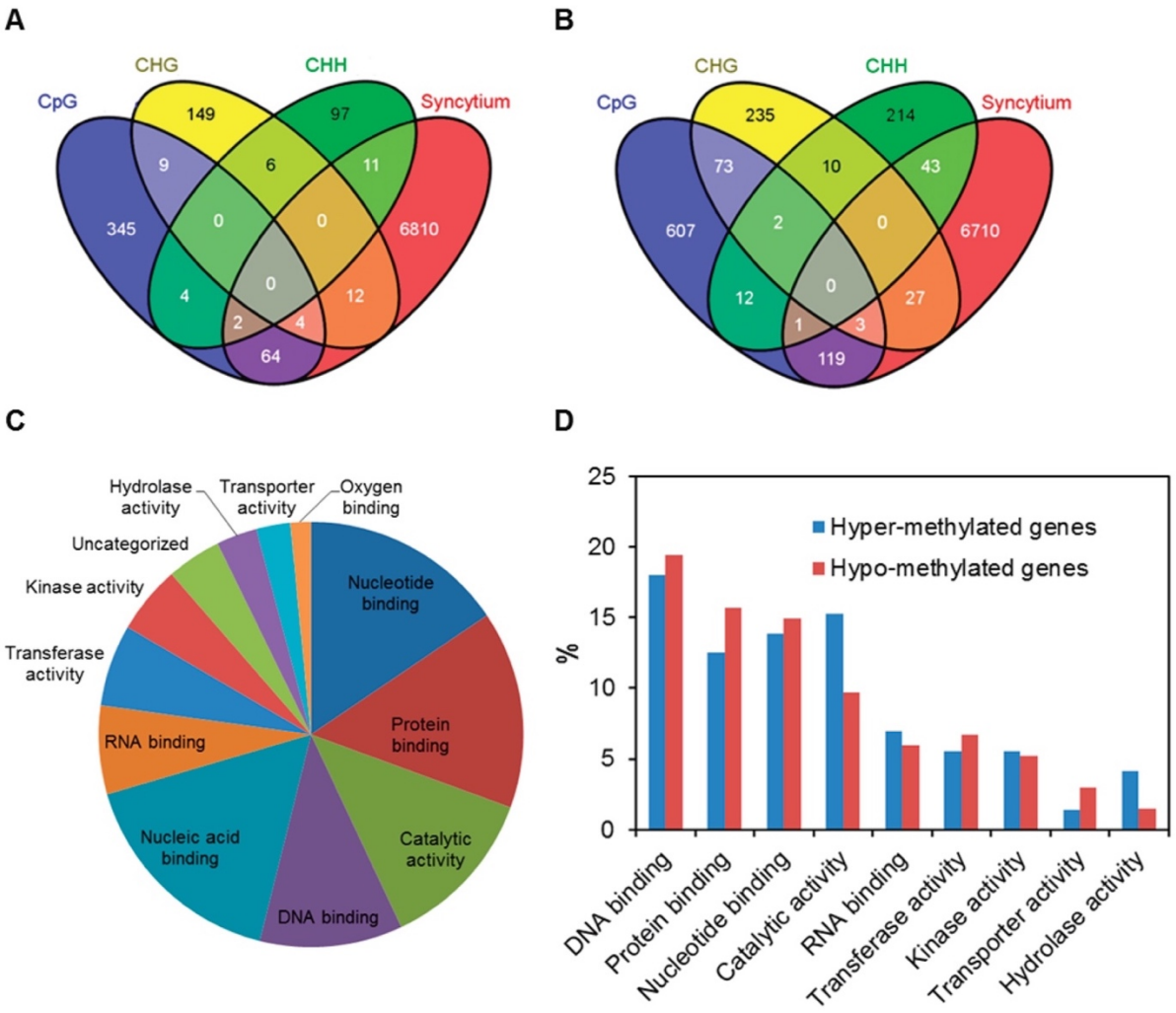

Figure 2.5 Identification of syncytium genes that are under methylation control. (A and B) Venn diagrams showing overlap between syncytium differentially expressed genes and those identified as hyper-methylated genes (A) or hypo-methylated genes (B) in various sequence contexts. Ninety-three of the differentially hyper-methylated genes and 193 of the differentially hypo-methylated genes were identified as overlapping with the 6,903 syncytium-differentially expressed genes. (C) Gene Ontology categorization of the molecular function of 278 unique syncytium-differentially methylated genes. Only categories represented by at least three genes were included. (D) Comparison of Gene Ontology categorization of the molecular function between the 93 hyper-methylated and the 193 hypo-methylated genes overlapping with syncytium differentially expressed genes.

Chapter 3: Methylome changes mediated by the soybean cyst nematode resistance gene *GmSHMT08*

Running Title: GmSHMT08-mediatied methylome changes

Title: Methylome changes mediated by the soybean cyst nematode resistance gene *GmSHMT08*

Aditi Rambani[1], Vince Pantalone[1], Songnan Yang[1], J. Hollis Rice[1], Prakash R. Arelli[2], Khalid Meksem[3], C. Neal Stewart[1], Tarek Hewezi[1*]

[1]Department of Plant Sciences, University of Tennessee, Knoxville, Tennessee 37996

[2]Crop Genetics Research Unit, USDA-ARS, Jackson, Tennessee, 38301

[3]Department of Plant, Soil, and Agricultural Systems, Southern Illinois University, Carbondale, Illinois 62901

To be submitted to Nature Genetics

I helped design experiments for this project. I prepared libraries for bisulfite and RNA sequencing. I analyzed data from methylome and transcriptome. I built a database with TE annotation, TE coordinates and TE distance from genes for the latest version of soybean assembly (Wm82.a2.v1). I contributed towards writing portion of the manuscript describing the methods I performed.

Abstract

Rhg4 is a major locus required for soybean cyst nematode (SCN) resistance in soybean accessions containing low copy of repeat unit at Rhg1 locus. *Rhg4* encodes a serine hydroxymethyltransferase (GmSHMT08), whose function in nematode resistance remains mostly unknown. Key role of GmSHMT08 in reprograming soybean methylome during SCN infection was elucidated using a pair of highly homozygous near-isogenic lines (NILs) containing the resistant or susceptible *Rhg4* allele. NILs had differential levels of methylation under non-infected conditions. The substantial significant differences in DNA methylation patterns between the NILs were associated with differential gene expression; this association may prime their response to SCN parasitism. In response to SCN infection, the NILs exhibited contrasted DNA methylation patterns with the methylome of the susceptible line being more dynamic than that of the resistant line. Comparing the methylomes of the parental lines with that of the NILs resulted in the identification of heritable as well as novel non-parental differentially methylated regions overlapping with genes related to SCN parasitism of soybean. Together, our analyses provide insights into the biochemical basis of *Rhg4* function in SCN resistance.

1. Introduction

DNA methylation is a consequential epigenetic factor that impacts gene expression, transposon mobility, genomic stability and imprinting (He et al., 2011). Previous studies identified key enzymes that carry-out cytosine DNA methylation in various sequence contexts. After DNA replication cytosine methylation in CG and CHG is maintained through the activity of METHYLTRANSFERASE1 (MET1) and CHROMOMETHYLASE3 (CMT3), respectively. Also, a small portion of CHH methylation can be maintained by CMT2 following DNA replication, but the large majority of CHH methylation sites are reestablished de novo. De novo DNA methylation in CG and non-CG contexts is carried-out through the synchronized activity of the RNA-directed DNA methylation (RdDM) pathway and DOMAINS REARRANGED METHYLTRANSFERASE 2 (Matzke and Mosher, 2014). Failure to faithfully maintain DNA methylation status by the maintenance enzymes after fertilization may cause spontaneous DNA methylation polymorphisms refer to as epialleles (Mosher et al., 2009; Slotkin et al., 2009; Schmitz et al., 2011). Epialleles can be induced by developmental and environmental stimuli as well as through transposon insertions and genome rearrangements that induce directed chromatin modifications (Hollister and Gaut, 2009; Taudt et al., 2016; Springer and Schmitz, 2017).

DNA methylation patterns of Arabidopsis lines generated through single-seed descent for 30 generations with their ancestral lines revealed that spontaneous loss and gain of DNA methylation at individual cytosine sites occur at high levels mainly in the genic regions in the CG sequence context (Becker et al., 2011; Schmitz et al., 2011). Regions with continuous cytosine methylation polymorphisms were also found but at relatively lower rate (Becker et al., 2011; Schmitz et al., 2011). Studies of inheritance and stability of DNA methylation patterns in maize and soybean using populations of recombinant inbred lines (RILs) provided additional evidences for transgenerational inheritance of DNA methylation variants over several generations in the segregating populations (Eichten et al., 2013; Regulski et al., 2013; Schmitz et al., 2013).

Similar to spontaneously generated DNA methylation variants experimentally induced DNA methylation changes may result in generation of novel non-parental DNA methylation polymorphisms that are heritable. For example, DNA hypomethylation induced by mutation in the Arabidopsis *DDM1* gene was stably transmitted over several generations (Kakutani et al., 1999). Similarly, *ddm1*-induced hypermethylation of *BONSAI* gene was consistently maintained in the *ddm1* after recurrent self-pollination (Saze and Kakutani, 2007). More comprehensive studies of transgenerational inheritance of DNA methylation variants were conducted in Arabidopsis using epigenetic recombinant inbred lines known as epiRILs that were generated by crossing the hypomethylated mutants *met1* or *ddm1* with wild-type plants (Johannes et al., 2009; Reinders et al., 2009). Analysis of DNA methylation profiles in these epiRILs documented the transgenerational inheritance of parental DNA methylation patterns in addition to the presence of newly acquired non-parental methylation variants (Johannes et al., 2009; Reinders et al., 2009). Furthermore, phenotypic analysis of the epiRILs for various traits including plant growth, plant height, flowering time and response to biotic and abiotic stresses revealed high degree of heritability (Johannes et al., 2009; Reinders et al., 2009) indicating that epiallelic variations may contribute to the heritability of complex traits. More recently, it has been shown that combining hypomethylated and normally methylated genomes in F1 plants triggers substantial reprogramming of plant methylomes that may result in novel and heritable epialleles (Rigal et al., 2016). Therefore, it is most plausible that interference with DNA methylation programs may spontaneously trigger heritable epigenetic variations that may be conditioned by *cis*- and/or *trans*-acting differences.

In soybean, soybean cyst nematode (SCN, *Heterodera glycines*) is the most damaging pathogen, causing significant yield and quality losses. SCN induces vascular root cells to fuse and form a permanent feeding structure, the syncytium, essential for nematode development and maturity. Resistance to SCN is conferred by two main loci, Rhg1 (for resistance to *H. glycines* and Rhg4, at chromosome 18 and 8, respectively. The Rhg1 locus contains three genes within a 31-kb repeat region that encode an α-SNAP protein, a putative amino acid transporter, and a wound-inducible protein (Cook et al., 2012).

Increased expression of these three genes, mediated by high copy number, was found to contribute to SCN resistance in an additive manner. Rhg4 locus contains only one gene encoding serine hydroxymethyltransferase (GmSHMT08) (Liu et al., 2012). SCN resistance in commercially available soybean cultivars is gained from two main sources that include plant introduction (PI) 88788 and Peking (Concibido et al., 2004). (PI) 88788-drived resistance is mediated by high copy number of rhg1-b allele (7 to 10 copies) (Cook et al., 2012), and cultivars produced from this source display slow degeneration of the nematode feeding site, leading to delayed arrest of nematode development (Kim et al., 2010). Peking-derived resistance requires both Rhg4 and rhg1-a allele (Meksem et al., 2001), and cultivars produced from this source display stronger and faster resistance response, leading to rapid arrest of nematode development at the infective juvenile stage (Mahalingam and Skorupska, 1996; Klink et al., 2009). However, molecular mechanisms through which Rhg1 and Rhg4 mediate SCN resistance remain to be elucidated.

SHMTs are key enzymes involved in one-carbon metabolism, a housekeeping cellular function that supports various physiological processes including redox defense and DNA methylation (Liu et al., 2012; Locasale, 2013; Mitchum, 2016).The anticipated function of Rhg4 (GmSHMT08) in redox defense is consistent with the activation of significant numbers of genes associated with oxidative stress, hypersensitive responses and programmed cell death in the syncytium formed in Peking (Klink et al., 2009). The potential implication of Rhg4 in modulating plant DNA methylation landscape is striking since recent studies indicated that level and pattern of plant DNA methylation are considerably modulated during cyst nematode infection (Rambani et al., 2015; Hewezi et al., 2017). In this study, the function of GmSHMT08 in establishing DNA methylome landscapes of soybean roots during SCN infection was studied using highly homozygous near-isogenic lines differing at Rhg4 locus. Our analyses provide unprecedented insights into the role of GmSHMT08 in reprograming soybean methylomes that may prime plant response to SCN parasitism.

2. Results

2.1 Developing near-isogenic soybean lines differing at the SCN resistance gene *Rhg4*

Two near isogenic lines (NILs), TN09-16 and TN09-29, respectively containing the susceptible and resistant allele of *Rhg4*, were generated and provided by Dr. Vince Pantalone at the[1]Department of Plant Sciences, University of Tennessee. These NILs are highly homozygous recombinant inbred lines derived from individual F13 generation single plants from a cross between the SCN-resistant variety Fowler and the SCN-susceptible variety Anand. Homozygosity level of these NILs was estimated to be 0.9998 based on the number of inbreeding generations after the creation of F1. SCN resistance in the parental line Fowler was acquired from the Plant Introduction (PI) 437654. PI 437654, which exhibit 'Peking-type' resistance, has been shown to carry the SCN resistance gene *Rhg4* and three copies of *Rhg1a* (Concibido et al., 2004; Cook et al., 2014). Initially, we used Simple Sequence Repeat (SSR) markers associated with *Rhg1* (Satt309) and *Rhg4* (Satt162 and Satt632) (Kazi et al., 2010) to test potential genetic differences between these two NILs at *Rhg1* and *Rhg4* loci. This analysis was performed in Dr. Prakash Arelli's laboratory at Crop Genetics Research Unit, USDA-ARS, Jackson, Tennessee. The resistant line TN09-29 inherited the two resistant alleles from its Hartwig ancestry. The susceptible line TN09-016, however, did not inherit the *Rhg4* allele but did receive the *Rhg1* resistant allele. We next cloned and sequenced the SCN resistance genes *soluble NSF attachment proteins* (*GmSNAP18*) and the *serine hydroxymethyltransferase* (*GmSHMT08*) at the *Rhg1* and *Rhg4* loci, respectively from the isogenic lines. While the nucleotide sequences of *GmSNAP18* were identical in both lines, *GmSHMT08* showed two single nucleotide polymorphisms (SNPs), leading to R130P and Y358N amino acid substitutions between the TN09-29 and TN09-16 (Figure S.3.1). These two amino acid substitutions were previously reported to establish the difference between the resistant and susceptible alleles of *Rhg4* (Liu et al., 2012).

Nematode susceptibility assays of the parental lines and the NILs were conducted against SCN HG Type 0 (race 3) in Dr. Prakash Arelli's laboratory at Crop Genetics Research Unit, USDA-ARS, Jackson, Tennessee. The parental line Anand showed very high level of SCN susceptibility with more than 100 cysts per plants. In contrast, Fowler showed high level of resistance with only one or two cysts can occasionally be seen in the inoculated plants (Figure S.3.2). Similarly, the NILs showed completely opposite responses to SCN infection. TN09-16 showed very high level of susceptibility, whereas TN09-29 showed complete resistance to SCN (Figure S.3.2). TN09-16 and TN09-29 were also evaluated for SCN resistance in the USDA Southern Uniform Soybean Test program (Gillen and Shelton, 2012). TN09-29 showed strong resistance to race 2 (HG Type 1.2.5.7), race 3, and race 5 (HG Type 2.5.7) with the highest possible resistant score ratings (1, 1, 1). In contrast, TN09-16 showed high level of susceptibility rated at scores of 4, 4, and 4, for SCN races 2, 3, and 5, respectively (Gillen and Shelton, 2012).These results are consistent with previous reports indicating that *Rhg4* is required for complete resistance against SCN in the PI 437654-derived germplasms (Meksem et al., 2001; Brucker et al., 2005).

2.2 The genomes of the near isogenic lines are significantly differentially methylated

Taking into consideration the function of GmSHMT08 in cellular methylation (Locasale, 2013; Mitchum, 2016), we therefore examined its potential impact on the methylome of the NILs under non-infected conditions. Seeds of the NILs TN09-16 and TN09-29 were germinated and root tissues of non-infected one-week-old seedlings were collected in three biological independent samples. DNA were isolated from these six samples and used to prepare methylC-seq libraries. The libraries were sequenced using Illumina HiSeq 2500 sequencing platform. We obtained 1.190E+09 100-bp reads for the susceptible TN09-16 line and 1.159E+09 reads for the resistant TN09-29 line, providing more than 100 X coverage of the soybean genomes. Bisulfite conversion rate was estimated using the λ phage genome and found to be greater than 99.7%. Differentially

methylated cytosines in the CG, CHH and CHG sequence contexts were considered for downstream analyses only if they covered by at least 10 high-quality reads. We compared the global methylation levels between TN09-16 and TN09-29 over genes and transposable elements (TEs) in the CG, CHG, and CHH contexts. Interestingly, differences in the global methylation levels between the two lines were detected in gene body as well as upstream and downstream regions in all sequence contexts (Figure 3.1 A-C). The susceptible line TN09-16 showed higher methylation levels than the resistant TN09-29 line over gene body and the flanking regions in the CHG and CHH contexts (Figure 3.1 B and C). In the CG context, TN09-16 showed higher methylation level in the flanking regions but lower methylation in gene body (Figure 3.1A). Differences in the global methylation levels between the two lines were also observed in TEs (Figure 3.1 D-F). TN09-16 showed higher methylation levels than TN09-29 in the TE flaking regions in all sequence contexts (Figure 3.1 D-F). Over the body of TEs, the global methylation levels of TN09-16 and TN09-29 in the CHG context were comparable (Figure 3.1 E). In CG and CHH, however, the differences between the two lines were noticeable. TN09-16 showed higher CHH-methylation and lower CG-methylation than TN09-29 (Figure 3.1 D and F). These initial analyses indicate that the genomes of these NILs are considerably differentially methylated.

To identify genomic regions with the most significant methylation differences, the genomes of these two lines were divided into 200-bp non-overlapping bins and then differentially methylated regions (DMRs) with at least 50% methylation differences were identified using a false discovery rate (FDR) of 0.01. Using these stringent criteria, 21,852 unique DMRs between TN09-16 and TN09-29 were identified. Of these 4,180 and 11,211 DMRs overlapped with protein-coding genes and TEs, respectively (Figure 3.2). The remaining DMRs are located in unannotated regions of the genome. As shown in Figure 3.2, the majority of these DMRs occurred in the CHG and CG contexts, with approximately two-thirds of these regions being hypermethylated in the susceptible line TN09-16 compared with the resistant line TN09-29. When the 4,180 DMRs were mapped to various annotated features of protein-coding genes we found that CG-DMRs occur mostly in gene body and to a much lesser extent in gene promoters and 5' and 3' untranslated regions (UTRs) (Figure S.3.3). Notably, 70% of CHG-DMRs were found in

gene body (Figure S.3.3). In contrast to CG- and CHG-DMRs, the number of CHH-DMRs overlapping with gene features was relatively small (Figure S.3.3). TE-associated DMRs were mapped to various transposon families. Remarkably, 88% of these DMRs were associated with long terminal repeat (LTR) retrotransposons (Figure S.3.4). DMRs overlapping with the DNA transposons Helitron and TIR constituted about 10% of the total number of TE-associated DMRs (Figure S.3.4).

Gene Ontology (GO) enrichment analysis of the 3,666 differentially methylated genes (DMGs) between TN09-16 and TN09-29 revealed overrepresentation of genes involved in various biological processes (Figure 3.3A). Of note is that ontologies with functions related to chromatin silencing, RNA interference, DNA repair, production of siRNAs, and histone H3-K9 methylation were significantly overrepresented. Thus, the enrichment of genes related to epigenetic modifications among the DMGs may explain the remarkable methylome differences between these two NILs.

2.3 Differential DNA methylation between the isogenic lines may prime their responses to SCN infection

To examine the degree to which cytosine methylation impacts gene expression in the isogenic lines, RNA-seq libraries were generated from the same root samples used for DNA methylation analysis. We identified 948 differentially expressed genes (DEGs) between TN09-16 and TN09-29 at a FDR of 0.05. GO term analysis revealed statistically significant enrichment for categories related to wounding response, defense response, membrane disassembly, and intracellular signal transduction (Figure 3.3B). This finding suggests that differences in priming response between TN09-16 and TN09-29 may contribute to their contrasted response to SCN infection. To examine this suggestion, we compared the 948 DEGs with our list of previously identified syncytium DEGs (6,903) and 267 genes were identified as common between these two gene lists. This significant overlap (28.16%, $\chi 2 = 38.91$, $P = 1.81E\text{-}08$) further supports our suggestion. We next compared the 948-gene list with the 3,666 DMGs to determine if there is a significant

enrichment for the DMGs among the DEGs. Fifty-two genes were common to both gene lists, revealing a significant enrichment (5.48%, $\chi 2 = 27.76$, $P = 4.07E-06$) for the DMGs among the DEGs (Figure 3.3C). The methylation patterns of these 52 genes seem to impact their expression levels. For example, promoter and gene body hypermethylation of Glyma.09G133900 in TN09-16 was associated with a significant gene downregulation (Figure 3.3D). In addition, we identified 9 DEGs that are located with 2 kb from differentially methylated TEs. Together, these data indicate that under non-infected conditions differential DNA methylation contributes to differential gene expression between the NILs that may prime their responses to SCN parasitism.

2.4 The susceptible and resistant isogenic lines exhibit contrasted DNA methylation patterns in response to SCN infection

We examined whether the methylomes of the TN09-16 and TN09-29 are similarly altered in response to SCN (race 3) infection. MethylC-seq libraries were constructed from SCN-infected roots at 5-day post infection (dpi) and compared with control samples. Both infected and non-infected libraries were prepared from root tissues collected from the same experiments at the same time. Differentially methylated cytosines were identified as described above and global methylation levels were compared between infected and non-infected samples over genes and TEs in all sequence contexts. In response to SCN infection, the susceptible line TN09-16 showed reduced methylation levels over protein-coding genes in all sequence contexts compared with non-infected control (Figure 3.4 A-C). In sharp contrast, the SCN-infected samples of the resistant line TN09-29 showed increased methylation levels over protein-coding genes in all sequence contexts compared with non-infected control (Figure 3.4 G-I).

Differences in global methylation patterns over TEs in response to SCN infection were also observed between the NILs. Infected TN09-16 samples showed reduced methylation levels over the body of TEs and flaking regions in all sequence contexts in comparison with the non-infected control samples (Figure 3.4 D-F). On the contrary,

infected TN09-29 samples showed increased methylation level over the body of TEs and flaking regions in CHH context relative to non-infected samples (Figure 3.4 L). In the CG and CHG contexts, however, increased methylation levels in the TN09-29–infected samples were observed only over the TE flaking regions (Figure 3.4 J and K). These analyses indicate that the NILs exhibit contrasted DNA methylation patterns in response to SCN infection with hypomethylation being more predominant in the susceptible line, whereas hypermethylation being more predominant in the resistant line.

2.5 DNA methylation patterns associated with the susceptible and resistant responses are highly specific

To localize the genome-wide DNA methylation profiles induced by SCN in both lines we identified DMRs as indicated above and mapped them to the annotated protein-coding genes and TEs. In response to SCN infection we identified 50,040 DMRs in TN09-016. Of these, 7,585 (15.16%) overlapped with protein-coding genes, and 28,100 (56.16%) overlapped with TEs (Figure 3.5 A). Of the 28,100 TE-associated DMRs 1,676 (5.96%) were located in genes. Notably, the number of DMRs in TN09-29 was dramatically lower (Figure 3.5 B). A total number of 5,080 DMRs were identified in the TN09-29-infected samples compared with controls (Figure 3.5B). Of these, 1,296 (25.51%) overlapped with protein-coding genes, and 2,356 (46.38%) overlapped with TEs (Figure 3.5B). Of the 2,356 TE-associated DMRs 178 (7.55%) were located in genes. When the DMRs overlapping with protein-coding genes or TEs were compared between the two lines, only 74 DMRs were common (Figure 3.5C), indicating that DNA methylation patterns associated with the susceptible and resistant responses are highly specific. This indication was further supported by GO term enrichment analysis showing the association of the DMGs in TN09-16 and TN09-29 with different GO biological process categories (Figure 3.5D and E).

To better understand the methylation differences between TN09-16 and TN09-29 in response to SCN infection, we examined the methylation level, direction, and sequence contexts of the DMRs. The numbers of hyper- and hypo-DMRs associated with protein-coding genes and TEs in the CG, CHG and CHH contexts are shown in Figure 3.5F-K. With the exception of the CG-DMRs overlapping with TEs (Figure 3.5I) the numbers of hypo-DMRs in various sequence contexts in TN09-16 were higher than hyper-DMRs. In TN09-29, however, we observed the opposite trend with hyper-DMRs being more predominant than hypo-DMRs in all cases except CHH-DMRs associated with TEs (Figure 3.5K). This trend is also evident when hyper- and hypo-DMRs overlapping with various annotated features of protein-coding genes (Figure S.3.5) and transposon families (Figure S.3.6) were compared within each line. Together, these data indicate that DNA methylation reprograming occurs prominently during the susceptible interaction and to a much lesser extent during the resistant interaction.

2.6 DNA methylation reprograming during the susceptible interaction contributes to gene expression changes

We further studied gene expression changes in the TN09-16 and TN09-29 in response to SCN using RNA-seq approach. The RNA libraries were prepared from the same tissue samples used for DNA methylation analysis to facilitate examining the potential link between DNA methylation and transcriptome changes. Because of the heterogeneity nature of SCN-infected roots we used a less stringent P value cutoff of < 0.05 and a FDR < 0.1 to identify DEGs. We identified 1,668 and 112 DEGs in TN09-16 and TN09-29, respectively at 5 d post SCN infection. The low number of the DEGs identified in the resistant lines may reflect the localized response to SCN infection compared with the susceptible line in which localized and systemic responses may occur both in the developing syncytium as well as in cells far from the infection sites. GO analysis revealed a significant enrichment of three biological process terms associated with plant responses to oxidative stress, chemical stimulus, and oxidation reduction among the TN09-29 DEGs. These results are consistent with the previous reports associating

oxidoreductase activity and oxidative stress response with Peking-type resistance (Klink et al., 2007; Klink et al., 2009; Klink et al., 2011), and support the potential function of GmSHMT08 in redox defense. Among the TN09-16 DEGs genes we noted a significant enrichment of biological process terms corresponding to plant response to stimulus and signaling of various phytohormones, including ethylene, salicylic acid, jasmonic acid, and abscisic acid (Figure 3.6A). Genes involved in ROS-mediated defense response, secondary cell wall biogenesis, cell wall loosening, membrane disassembly, and cellular responses to wounding and chitin were also overrepresented among the TN09-16 DEGs (Figure 3.6A).

Next, we determined the potential association between DNA methylation and gene expression changes. We compared the 112 DEGs and the 1293 DMGs identified in the TN09-29 upon SCN infection and only one gene (Glyma.04G180400) was common between the two gene lists (Figure 3.6B). Also, we examined if differentially methylated TEs are located within 2 kb upstream or downstream of the 112 DEGs. Only one gene (Glyma.06G102000) was found to contain a differentially methylated TE 1.813 kb from its transcriptional start site (TSS) (Figure 3.6B). These results imply that differential DNA methylation doesn't seem to directly impact gene transcription during the resistant interaction. Similarly, we compared the 1668 DEGs and the 6252 DMGs identified in the TN09-16 after SCN infection and 123 genes were common between the two gene lists (Figure 3.6C), implying that DMGs are statistically significantly enriched among the DEGs (7.37%, $\chi 2$ = 140.3, P = 3.27E-30). Furthermore, we identified 50 DEGs containing differentially methylated TEs in their gene body or promoters, 2kb upstream of the TSS, resulting in a unique list of 147 differentially expressed DMGs (Figure 3.6C). Of these 147 genes, 47 have been previously shown to change expression in the SCN-induced syncytium, providing additional support for the involvement of these genes in plant-SCN interaction. Thus, we conclude that unlike the resistant interaction, DNA methylation reprograming during the susceptible interaction may directly contribute to gene expression changes.

2.7 Identification of stably inherited DMRs potentially associated with SCN infection

To identify stably inherited DMRs in the genic regions with potential association with SCN resistance/susceptibility we searched for differential methylation in the isogenic lines that are inherited from the parents. In other words, we were looking for DMRs with the exact genomic coordinates that are hypermethylated in the susceptible parent (Anand) and the susceptible line TN09-16 but hypomethylated in the resistant parent (Fowler) and the resistant line TN09-29, and vice versa (hypomethylated in Anand and TN09-16 but hypermethylated in Fowler and TN09-29). Therefore, we generated methylC-seq libraries from the two parental lines Fowler and Anand using non-infected root tissues collected from the same experimental settings described above. Differentially methylated cytosines were identified and global methylation levels over genes and TEs in all sequence contexts were compared between the parental lines. Interestingly, the susceptible parent (Anand) showed higher methylation levels than the resistant parent (Fowler) over genes and TEs in all sequence contexts (Figure S.3.7). A significant number of DMRs between the parental lines were also detected. Of the 45,603 DMRs detected, 7,000 mapped to protein-coding genes and 21,667 mapped to TEs. Consistent with global methylation patterns, the majority (65.70 %) of DMRs were hypermethylated in the susceptible parent (Anand) compared with the resistant parent (Fowler). These DNA methylation patterns are consistent with our results mentioned above and showing increased global methylation levels and hyper-DMRs in the susceptible line TN09-16 compared to the resistant line TN09-29.

We then compared the methylomes of the parental lines and that of the isogenic lines. We identified 59 DMRs in the isogenic lines with differential methylation patterns that were inherited from the parents (Figure 3.7A). As shown in Figure 3.7A, the 59 DMRs are grouped into two main clusters. The first cluster contains 38 regions that were hypomethylated in Fowler and TN09-29 but hypermethylated in Anand and TN09-016. An example of these regions is highlighted in Figure 3.7B. The second cluster contains 21 regions that were hypermethylated in Fowler and TN09-29 but hypomethylated in

Anand and TN09-16 (Figure 3.7A). These regions showed differential methylation in the CG (36) and non-CG contexts (23) and are located in the gene body (38), promoter (17) and 5'UTR (4) (Figure 3.7C and D). These 59 DMRs overlapped with 59 unique protein-coding genes, four of them were previously reported as differentially expressed in soybean syncytium.

2.8 Identification of novel non-parental DMRs specific to the isogenic lines

We further extended these analyses by comparing the methylomes of the parental lines and the isogenic lines under non-infected conditions in order to identify novel non-parental DMRs unique to TN09-16 or TN09-29. Interestingly, 56 DMRs unique to TN09-16 were identified. Gain or loss of DNA methylation in these regions occurred to a similar extent and was opposite of that detected in the parental lines and TN09-29 (Figure 3.8A). An example of these regions is provided in Figure 3.8B. The opposite methylation patterns were detected in CG (37), non-CG (16), and both CG and non-CG contexts (3) (Figure 3.8C). The DMRs were located in gene body (42), gene promoters (9), and UTRs (5) (Figure 3.8D). The 56 DMRs overlapped with 55 protein-coding genes. This gene list included 9 of the previously identified syncytium DEGs, implying a role in the susceptible soybean-SCN interaction. We also identified a homolog of Arabidopsis mRNA splicing factor among this gene list. This finding promoted us to examine our RNA-seq data for differential usage of exon and exon-exon junctions using JunctionSeq package (Hartley and Mullikin, 2016). Using a FDR cut-off of 0.1, we identified 12 alternatively spliced genes with 13 differentially used exons/junctions when the transcriptomes of TN09-16 and TN09-29 were compared under non-infected conditions (Figure 3.8E). Of these exons/junctions 10 were significantly highly used in TN09-16 as in the case of Glyma17G149600 (Figure 3.8F), whereas the remaining 3 were significantly highly used in TN09-29. Interestingly, we found that 4 out of these 12 differentially spliced genes are among the syncytium DEGs, suggesting that novel non-parental hypomethylation of a splicing factor-encoding gene in TN09-16 may affect

alternative splicing of syncytium DEGs. Because DNA methylation can impact splicing efficiency (Regulski et al., 2013; Wang et al., 2016), we intersected these 12 genes with our 3,666 genes showing differential DNA methylation between TN09-16 and TN09-29. None of these genes was common to both lists, indicating that alternative splicing of these genes occurred independently of their methylation patterns.

Similarly, we compared the methylomes of the parental lines and the isogenic lines under non-infected conditions in order to identify novel non-parental DMRs unique to TN09-29. Interestingly, 106 DMRs specific to TN09-29 were identified (Figure 3.9A). Hyper- and hypomethylation in these 106 regions occurred to a similar extent and was opposite of that detected in the parental lines and TN09-16 (Figure 3.9A). An example of hypermethylated region specific to TN09-29 is provided in Figure 3.9B. Differential methylation of these DMRs was found in CG (61) and non-CG (45) (Figure 3.9C). The opposite methylation patterns were detected in gene promoter/5'UTR (28) and gene body/3'UTR (81) (Figure 3.9D). The 106 DMRs overlapped with 107 protein-coding genes, 11 of them previously reported to change expression in SCN-induced syncytium (Figure 3.9E and F). Together, these data suggest that novel non-parental DMRs specific to the isogenic lines may impact gene expression in the nematode feeding sites.

3. Discussion

Our analysis of the methylome landscapes mediated by the SCN resistance gene *Rhg4* was facilitated by using NILs differing in *Rhg4* locus and show opposite responses to SCN infection. A key role of *Rhg4* in establishing the root methylomes of the isogenic lines was demonstrated by the finding that the genomes of the isogenic lines were substantially differentially methylated both under non-infected and SCN-infected conditions. *Rhg4* is believed to be the key factor mediating these differences since no genes involved in DNA methylation machinery was identified as differentially expressed when we compared the root transcriptome of the isogenic lines. Our data also indicate that methylome differences in the isogenic lines were established de novo since only 59

genic regions showing opposite methylation patterns in these lines were inherited from the parents.

Under non-infected conditions, we observed a general trend of increased global DNA methylation in the susceptible line relative to the resistant line, specifically in the CHG and CHH contexts over protein-coding genes and TEs. However, the resistant line exhibited increased DNA methylation over the body of protein-coding genes and TEs in the CG context. This finding suggests that global decrease in CHG and CHH methylation in the resistant line was complemented with an increase of CG methylation. Unlike other plant species in which gene body methylation occurs exclusively at CG sites (Lister et al., 2008; Li et al., 2012; Regulski et al., 2013; Wang et al., 2016), our analysis revealed a significant number of CHG-DMRs in the body of protein-coding genes when the methylomes of the two isogenic lines were compared under non-infected conditions. This could be the results of high insertion rate of short TEs in the introns of these genes. However, we found that only 32% of the body CHG-methylated genes possessing DMTEs inside their transcribed regions. Another possibility is that CHG methylation in gene body may be linked to the activity of various histone demethylases that function in eliminating methylation at histone H3 lysine 9 (H3K9me). In Arabidopsis, it has been shown that loss-of-function of the H3K9 demethylase *IBMI* (Increase in BONSAI Methylation1) induced CHG methylation in the body of thousands of highly expressed genes (Saze et al., 2008; Miura et al., 2009; Inagaki et al., 2010). Thus, differential accumulation of H3K9me in the body of actively transcribed genes may be responsible for CMT3-mediated differential genic CHG methylation between the isogenic lines. We also examined if CHG-methylation in gene body was mechanistically associated with CG or CHH methylation. Interestingly, only 24 and 3% of body CHG-methylated genes showed differential methylation in CG or CHH contexts, respectively. This finding suggests that CHG methylation in transcribed regions is mostly independent of CG and CHH methylation. The isogenic lines also exhibited high level of differential DNA methylation in TEs, particularly LTR-type in the CHG context. This result may reflect differences in the transposition dynamics of Copia and Gypsy retrotransposons, which tend to be more vulnerable to DNA methylation than other types of TEs (Wang et al., 2015; Hewezi et al., 2017).

Rhg4-mediated methylome changes in the isogenic lines under non-infected conditions appeared to significantly impact the methylation patterns of several genes involved in epigenetic regulation. It is well known that various epigenetic components including DNA methylation, histone modification and siRNA accumulation are highly interconnected (Zilberman et al., 2008; Saze et al., 2012; Du et al., 2015). Thus, differential methylation of key genes involved in these pathways may be part of epigenetic feedback regulatory mechanisms that maintain epigenetic information. Maintenance of epigenetic information could contribute to priming of plant defense responses (Holoch and Moazed, 2015; To et al., 2015). Our RNA-seq analysis of the NILs under non-infected conditions pointed into a possible difference in some type of defense priming between the lines. This possibility was further reinforced, first, by our finding that the DEGs were highly enriched for genes involved in defense responses and biological processes associated with nematode susceptibility. Second, about one-third of the DEGs were previously shown to change expression in SCN-induced syncytium. Third, the DEGs included numerous marker genes of defense priming, including several WRKY transcription factors, ROS-related genes, and lipoxygenases (Martinez-Medina et al., 2016; Mauch-Mani et al., 2017). Recently, it has been shown that Arabidopsis hyper- and hypomethylated mutants developed opposite responses to pathogen infection that were dependent on the mutants' ability to prime salicylic acid-mediated defense responses and callose deposition (López Sánchez et al., 2016). In this context, DNA methylation can prime defense responses in a way that alters chromatin structure to expedite gene transcription (Martinez-Medina et al., 2016; Mauch-Mani et al., 2017).

The role of GmSHMT08 in establishing the methylomes of the isogenic lines became more evident when the methylomes of isogenic lines were determined under SCN-infected conditions. Consistent with the role of SHMTs in cellular methylation (Locasale, 2013; Mitchum, 2016), the susceptible line, which contains a non-functional allele of RHg4, exhibited reduced global methylation levels in both protein-coding genes and TEs, whereas the resistant line showed the opposite response of increased global methylation levels. This trend was observed in all three methylation contexts, suggesting that the SHMT may have significant impact on the whole methylation pathway. Also,

our results show that the methylome of the susceptible line is more dynamic than that of the resistant line in response to SCN infection, as 50,040 DMRs were identified in the susceptible line compared to only 5,080 DMRs in the resistant line. This dynamic can be explained by the many cellular processes that accompany syncytium formation and development during the susceptible interaction compared to localized cell death that occurs during the resistant interaction. In addition to the dramatic differences between the isogenic lines in term of methylation level and direction, differential DNA methylation patterns within protein-coding genes and TEs demonstrated remarkable level of specificity as only 74 DMRs were found common to both lines.

Despite the established association between DNA methylation and gene transcription (Niederhuth and Schmitz, 2017), our analysis revealed low overlap between the DEGs and the DMGs in the resistant line during SCN infection. This may be due to the dilution of localized gene expression changes at the infection sites by using whole roots in our analysis. Because DNA methylation can precede gene expression changes during cyst nematode infection (Hewezi et al., 2017), the DMGs may associate with gene transcription at later stage of infection. Another possibility is that DNA methylation patterns may regulate the steady-state expression of these genes, preventing their induction or repression during the resistant response. A role of DNA methylation as a secondary stabilizer of gene expression has been recently proposed (Smith and Meissner, 2013). Also, DNA methylation function mutually with other epigenetic modification (Zilberman et al., 2008; Du et al., 2015) and hence, one can anticipate that methylation status of these genes may necessitate additional epigenetic marks to influence gene transcription to the level of significance.

In contrast to the resistant interaction, DNA methylation reprograming during the susceptible interaction seem to directly impact gene transcription levels. A set of 147 differentially expressed/DMGs were identified in our analysis. Differential DNA methylation seems to impact cellular functions that are directly modulated by cyst nematode effectors. For example, genes involved in pectin demethylesterification and polyamine oxidation, which are targeted by the cyst nematode effectors cellulose binding protein and 10A06, respectively (Hewezi et al., 2008; Hewezi et al., 2010) were

among the 147 differentially expressed/DMGs. The potential effect of DNA methylation on gene transcription was obvious in many situations as in the case of an adaptin family protein gene, which was hypomethylated in the promoter, gene body, and 3'UTR in various sequence contexts and highly induced in response to SCN infection. Adaptin family proteins are involved in intracellular protein trafficking (Boehm and Bonifacino, 2001) and thus hypomethylation-mediated upregulation of this gene may facilitate assimilate flow to the syncytium. Another example of the impact of DNA methylation is the downregulation of cycling DOF factor 2 (CDF2), which was hypomethylated in gene body in the CG context, but was hypermethylated in the promoter region in the CHH context. In Arabidopsis, CDF2 has been reported to regulate the expression of a number of miRNA genes at both transcriptional and posttranscriptional levels by direct binding to miRNA promoters or through modulation of DCL1-mediated processing of primary miRNA transcripts (Sun et al., 2015). Thus, reprogramming of DNA methylation patterns may function in concert with other epigenetic pathways during SCN parasitism.

Increased activity of metabolism pathways is known to play central role in successful nematode parasitism (Hofmann et al., 2010). Our data indicate that DNA methylation contributes to the regulation of the transcriptional activity of several key primary metabolism genes specifically those associated with the metabolic processes of carbohydrate, glucan and malate, presumably to maintain metabolite levels at an active physiologic status compatible with nematode feeding and development. Associations between hyper- and hypomethylation and significant changes in gene expression were also observed for several genes involved in the biogenesis of primary and secondary cell walls, organization of actin and microtubules, defense responses, and signal transduction. These findings are in agreement with our previous reports showing the enrichment of genes encoding these functions among the DMGs induced in response to infection by cyst nematodes in soybean and Arabidopsis (Rambani et al., 2015; Hewezi et al., 2017). Thus, cyst nematode-induced differential methylation during the susceptible interaction appears to regulate similar cellular processes in various plant species.

The methylome analysis of the parental lines and the two isogenic lines allowed us to identify 56 and 106 regions that exhibit novel non-parental methylation patterns unique to TN09-16 and TN09-29, respectively. Gain or loss of DNA methylation in these regions seem to occur to a similar extent with preference observed for CG and CHG contexts. Gain of DNA methylation appears to be introduced in the isogenic lines, whether environmentally or genetically, during the 13 generations of breeding and was faithfully maintained through the activity of MET1 and CMT3. In contrast, loss of DNA methylation in certain genic regions could be the result of absence of corrective DNA methylation mechanisms that can restore DNA methylation. Failure to reestablish DNA methylation in specific genic regions may be owing to the loss of the repressive histone mark H3K9me2, which has been shown to be directly or indirectly linked to the activity of MET1 and CMT3-mediated DNA methylation in CG and non-CG contexts (Bernatavichute et al., 2008; Inagaki et al., 2010; Du et al., 2012).

The non-parental methylation patterns in these genic regions could be induced by transposons located in proximity to these DMR-associated genes, specifically those transposons that are prone to DNA methylation change under infected conditions. In agreement with this view we found that 12 of the 56 DMRs that are unique to the susceptible line TN09-16 were located within 2 Kb of differentially methylated TEs identified in TN09-16 in response to SCN infection. Previously, it has been shown that TEs can induce heritable epialleles by bringing neighboring genes under their own regulation (Slotkin and Martienssen, 2007). We also examined whether these 56 DMRs are particularly vulnerable to DNA methylation changes under SCN infected conditions. Interestingly, we found that 18 of these DMRs are among those identified in TN09-16 in response to SCN infection. Similarly, 20 of the 106 DMRs that are unique to TN09-29 are among those showing differential methylation under SCN infection. Together, these findings suggest that the non-parental methylation patterns occur in regions that are vulnerable to methylation changes and that some introduced variations in DNA methylation pattern can be inherited and stably transmitted to offspring. Recently, it has been reported that a small portion of stress-induced DNA methylation changes can be faithfully transmitted to next generations (Wibowo et al., 2016). The biological significance of the newly acquired DNA methylation patterns in these genes remains

elusive. However, the association of these regions with genes previously shown to change expression in the syncytium highlight sheds lights into a possible role of these in SCN parasitism of soybean.

4. Materials and Methods

4.1 Developing near-isogenic soybean lines differing in GmSHMT08 gene

Two NILs, TN09-16 and TN09-29, which exhibit susceptible and resistant responses, respectively, to SCN HG type 0 (race 3) were generated and provided by Dr. Vince Pantalone at the1Department of Plant Sciences, University of Tennessee. These NILs are highly homozygous recombinant inbred lines derived from individual F13 generation single plants from the cross 'Fowler' × 'Anand'. Anand was developed from the cross Holladay × Hartwig by the Missouri Agricultural Experiment Station and released in 1999 (Anand et al., 2001). Fowler was developed from the cross Hartwig × Holladay by the USDA-ARS at Jackson, TN, and released in 1999 (Young LD. 2001). Hartwig derived its resistance from Plant Introduction (PI) 437654. Simple sequence repeat (SSR) markers associated with *rhg1* (Satt309) and *Rhg4* (Satt162 and Satt632) (Kazi et al., 2010) were initially used to examine the genetic differences between these two NILs. In addition, the SCN resistance genes soluble NSF attachment proteins (GmSNAP18) and the serine hydroxymethyltransferase (GmSHMT08) at the *rhg1-b* and *Rhg4* loci, respectively, were amplified, ligated into pGMT-easy vector (Promega), and sequenced to further confirm the genetic differences between the NILs at GmSHMT08 locus.

4.2 Nematode susceptibility assays

Nematode susceptibility assays of isogenic lines and their parental lines were conducted in the greenhouse using SCN HG type 0 (race 3). Seeds of each line were planted in pots

(2 seeds per pot) containing soil : sand (1:1) mixture and organized in a randomized complete block design. Each pot was inoculated with approximately 4,000 eggs at seeding. Approximately 5 weeks after planting the cysts were blasted off the roots using high pressure water and counted under the microscope. Statistically significant differences between the parent as well as the isogenic lines and were calculated using t-tests with P value < 0.001.

4.3 Nematode inoculation and collection of root tissues

Soybean seeds of TN09-16 and TN09-29 were washed for 30 minutes under running water and then soaked in 10% bleach for 10 minutes. After this surface sterilization the seeds were washed again for 30 minutes to remove bleach remnants. The seeds were then germinated on wet germination paper in dark at 26 degrees for three days. Healthy looking 3 day old seedlings were selected for nematode inoculation. Freshly hatched second stage juveniles (J2s) of SCN HG type 0 (race 3) were surface-sterilized and then suspended in 0.1% sterile agarose solution at a concentration of approximately 500 J2s per 100 µL. Each seedling was inoculated with about 3000 J2s, by spreading the nematodes across the whole root of a seedling. Control plants were set up in exact same way, except mock inoculations were performed using 0.1% (w/v) agarose per seedling. Control and inoculated plants were arranged in replicates, each containing at least six plants and maintained in a controlled plant growth chamber at 26°C with 16-h light/8-h dark conditions as previously described (Rambani et al., 2015). Five days post SCN-inoculation, roots tissues were collected from both inoculated and non-inoculated soybean roots in three biologically independent replicates resulting in a total of 12 samples. Successful infection of each replicate was confirmed by examining one-fourth of the infected seedlings using acid fuschin stain (Daykin and Hussey, 1985). The two parental lines (Fowler and Anand) were only mock-inoculated in the same experimental sittings and a total of 6 biologically independent samples were similarly collected five days later. DNA and RNA were isolated from each sample and used construct methylC-seq and RNA-seq library.

4.4 Preparation of methylC-seq libraries

Genomic DNA of the infected and non-infected root samples was extracted using DNeasy Plant Mini Kit (Qiagen). Whole genome methylC-seq libraries were constructed as per protocol from Illumina TruSeq Library Prep kit (Illumina, San Diego, CA) with slight modifications of the bisulfite treatment. Briefly, about 2 µg of genomic DNA (gDNA) in addition to unmethylated lambda DNA were fragmented using Bioruptor (Diagenode Inc. USA, Denville, NJ) and then spiked with unmethylated fragmented lambda DNA (Promega, Madison, WI) that constitutes up to 2% of total concentration. Fragment size distribution of sheared DNA was verified using the Agilent Bioanalyzer 1000 DNA chip (Agilent Technologies, Santa Clara,CA). DNA fragments were then ligated to cytosine-methylated adapters (provided by Illumina) and then subjected to sodium bisulfite treatment using MethylCode™ Bisulfite Conversion Kit (Invitrogen, Grand Island, NY). DNA fragments between 400 and 500 bp were selected on the Pippin Prep system (Sage Sciences) and enriched by 10 cycles of PCR as recently described (Rambani et al., 2015). PCR products were then purified using Agencourt AMPure XP beads (Beckman Coulter, Inc., Brea, CA) and subsequently enriched using 5 additional PCR cycles according to Illumina's protocol. The PCR products were purified once more and library size distribution was examined using the Agilent Bioanalyzer 1000 DNA chip. Finally, the libraries were quantified and sequenced using Illumina HiSeq 2500 platform.

4.5 Identification of differentially methylated regions and overlapping genomic regions

Sequencing adapters were trimmed from bisulfite sequencing reads (BS reads) and low quality reads below phred threshold of 33 were removed using Trimmomatic (Bolger et al., 2014).Then high quality paired-end reads were aligned to the soybean reference

genome (Wm82.a2.v1) using bioinfomatic software Bismark with default parameters (Krueger and Andrews, 2011). Alignment files generated by Bismark were analyzed by the R bioconductor package methylKit (Akalin et al., 2012) to identify differentially methylated cytosines. Methylation status at each cytosine covered by at least 10 reads in the CG, CHG and CHH sequence contexts were calculated. A non-overlapping sliding window of 200 bp overall the 20 soybean chromosomes was used to identify DMRs with methylation difference of at least 50%. Significance of differentially hyper- and hypo-DMRs was determined using *q*-value less 1% as previously described (Rambani et al., 2015). DMRs were mapped to various genic regions including prompter (1 kb upstream of the transcription start site), 5' and 3' untranslated regions (UTRs), and gene body (transcribed region) using Bioconductor package rtracklayer in a custom R script (Lawrence et al., 2009). Methylation cytosine report files generated by Bismark were used to visualize global methylation levels over protein-coding genes and TEs using ViewBS package (https://github.com/xie186/ViewBS).

The most recent assembly of soybean genome (Wm82.a2.v1) was released without annotation of TEs. Thus, sequences of previously annotated and known TEs in soybean assembly were obtained from SoyTEdb (https://www.soybase.org/soytedb/). The TE sequences were used to mask the new soybean assembly using RepeatMasker (Smit et al., 2015) and output file with TE coordinates in the new assembly was created. Then, overlaps of DMRs with TEs belonging to various families were reported. BED tools (Quinlan and Hall, 2010) were used to calculate the distance from DMR-associated TEs to the nearest gene.

4.6 RNA library preparation and transcriptome analysis

mRNA was isolated using magnetic mRNA isolation kit (NEB). NEBnext mRNA library prep master mix (NEB) was used to build libraries following manufacturer's protocol. RNA-seq libraries were sequenced on Illumina Hiseq 2500 platform. Quality of paired-ended reads was verified with FastQC (version 0.11.4)

(http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Adapter sequences and low-quality reads were removed using Trimmomatic (version 0.35) (Bolger et al., 2014). Qualified reads were then mapped to the soybean reference genome (Wm82.a2.v1) using TOPHAT v.2.0.13 with default parameters (Trapnell et al., 2009). Reads mapped to multiple loci were discarded and numbers of uniquely mapped reads per gene were determined using HTSeq (Anders et al., 2015). Counts generated by HTSeq were used to determine differentially expressed genes using the R bioconductor package edgeR (Robinson et al., 2010). Genes with false discovery rate less than 0.1 or 0.05 were considered significantly differentially expressed. Separate count files were generated using the python-based package QoRTs for counting sequencing reads spanning exons for every gene (Hartley and Mullikin, 2015).The count files were used with the bioconductor R package JunctionSeq to determine differentially spliced transcripts (Hartley and Mullikin, 2016).

4.7 GO terms enrichment analysis

GO terms enrichment analysis of differentially methylated genes and differentially expressed genes were determined using soybase tools and AgriGO database (Du et al., 2010). Statistically significant enriched GO terms were calculated using Fisher's exact test and Bonferroni multi-test adjustment with a q value less than 0.05. GO terms were clustered based on semantic similarity to other GO terms in Uniprot database using REVIGO (Supek et al., 2011).

References

Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, Mason CE (2012) methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome biology **13:** R87

Anders S, Pyl PT, Huber W (2015) HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics **31:** 166-169

Becker C, Hagmann J, Müller J, Koenig D, Stegle O, Borgwardt K, Weigel D (2011) Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. Nature **480:** 245-249

Bernatavichute YV, Zhang X, Cokus S, Pellegrini M, Jacobsen SE (2008) Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in *Arabidopsis thaliana*. PloS one **3:** e3156

Boehm M, Bonifacino JS (2001) Adaptins the final recount. Molecular biology of the cell **12:** 2907-2920

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30:** 2114-2120

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

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

Cook DE, Bayless AM, Wang K, Guo X, Song Q, Jiang J, Bent AF (2014) Distinct copy number, coding sequence, and locus methylation patterns underlie Rhg1-mediated soybean resistance to soybean cyst nematode. Plant physiology **165:** 630-647

Cook DE, Lee TG, Guo X, Melito S, Wang K, Bayless AM, Wang J, Hughes TJ, Willis DK, Clemente TE (2012) Copy number variation of multiple genes at Rhg1 mediates nematode resistance in soybean. Science **338:** 1206-1209

Daykin M, Hussey R (1985) Staining and histopathological techniques in nematology.

Du J, Johnson LM, Jacobsen SE, Patel DJ (2015) DNA methylation pathways and their crosstalk with histone methylation. Nature reviews molecular cell biology **16:** 519-532

Du J, Zhong X, Bernatavichute YV, Stroud H, Feng S, Caro E, Vashisht AA, Terragni J, Chin HG, Tu A (2012) Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. Cell **151:** 167-180

Du Z, Zhou X, Ling Y, Zhang Z, Su Z (2010) agriGO: a GO analysis toolkit for the agricultural community. Nucleic acids research **38:** W64-W70

Eichten SR, Briskine R, Song J, Li Q, Swanson-Wagner R, Hermanson PJ, Waters AJ, Starr E, West PT, Tiffin P (2013) Epigenetic and genetic influences on DNA methylation variation in maize populations. The Plant Cell **25:** 2783-2797

Gillen A, Shelton G (2012) Uniform soybean tests, southern states, 2011. USDA–ARS, Stoneville, MS. ht tp. *In*,

Hartley SW, Mullikin JC (2015) QoRTs: a comprehensive toolset for quality control and data processing of RNA-Seq experiments. BMC bioinformatics **16:** 224

Hartley SW, Mullikin JC (2016) Detection and visualization of differential splicing in RNA-Seq data with JunctionSeq. Nucleic acids research **44:** e127-e127

He X-J, Chen T, Zhu J-K (2011) Regulation and function of DNA methylation in plants and animals. Cell research **21:** 442-465

Hewezi T, Howe P, Maier TR, Hussey RS, Mitchum MG, Davis EL, Baum TJ (2008) Cellulose binding protein from the parasitic nematode *Heterodera schachtii* interacts with Arabidopsis pectin methylesterase: cooperative cell wall modification during parasitism. The Plant Cell **20:** 3080-3093

Hewezi T, Howe PJ, Maier TR, Hussey RS, Mitchum MG, Davis EL, Baum TJ (2010) Arabidopsis spermidine synthase is targeted by an effector protein of the cyst nematode *Heterodera schachtii*. Plant physiology **152:** 968-984

Hewezi T, Lane T, Piya S, Rambani A, Rice JH, Staton M (2017) Cyst nematode parasitism induces dynamic changes in the root epigenome. Plant Physiology**:** pp. 01948.02016

Hofmann J, El Ashry AEN, Anwar S, Erban A, Kopka J, Grundler F (2010) Metabolic profiling reveals local and systemic responses of host plants to nematode parasitism. The Plant Journal **62:** 1058-1071

Hollister JD, Gaut BS (2009) Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome research **19:** 1419-1428

Holoch D, Moazed D (2015) RNA-mediated epigenetic regulation of gene expression. Nature Reviews Genetics **16:** 71-84

Inagaki S, Miura-Kamio A, Nakamura Y, Lu F, Cui X, Cao X, Kimura H, Saze H, Kakutani T (2010) Autocatalytic differentiation of epigenetic modifications within the Arabidopsis genome. The EMBO journal **29:** 3496-3506

Johannes F, Porcher E, Teixeira FK, Saliba-Colombani V, Simon M, Agier N, Bulski A, Albuisson J, Heredia F, Audigier P (2009) Assessing the impact of transgenerational epigenetic variation on complex traits. PLoS genetics **5:** e1000530

Kakutani T, Munakata K, Richards EJ, Hirochika H (1999) Meiotically and mitotically stable inheritance of DNA hypomethylation induced by ddm1 mutation of *Arabidopsis thaliana*. Genetics **151:** 831-838

Kazi S, Shultz J, Afzal J, Hashmi R, Jasim M, Bond J, Arelli PR, Lightfoot DA (2010) Iso-lines and inbred-lines confirmed loci that underlie resistance from cultivar 'Hartwig'to three soybean cyst nematode populations. Theoretical and applied genetics **120:** 633-644

Kim Y-H, Kim K-S, Riggs RD (2010) Differential subcellular responses in resistance soybeans infected with soybean cyst nematode races. The Plant Pathology Journal **26:** 154-158

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

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

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

Krueger F, Andrews SR (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. bioinformatics **27:** 1571-1572

Lawrence M, Gentleman R, Carey V (2009) rtracklayer: an R package for interfacing with genome browsers. Bioinformatics **25:** 1841-1842

Li X, Zhu J, Hu F, Ge S, Ye M, Xiang H, Zhang G, Zheng X, Zhang H, Zhang S (2012) Single-base resolution maps of cultivated and wild rice methylomes and regulatory roles of DNA methylation in plant gene expression. BMC genomics **13:** 300

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell **133:** 523-536

Liu S, Kandoth PK, Warren SD, Yeckel G, Heinz R, Alden J, Yang C, Jamai A, El-Mellouki T, Juvale PS (2012) A soybean cyst nematode resistance gene points to a new mechanism of plant resistance to pathogens. Nature **492:** 256-260

Locasale JW (2013) Serine, glycine and one-carbon units: cancer metabolism in full circle. Nature reviews Cancer **13:** 572-583

López Sánchez A, Stassen JH, Furci L, Smith LM, Ton J (2016) The role of DNA (de) methylation in immune responsiveness of Arabidopsis. The Plant Journal **88:** 361-374

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

Martinez-Medina A, Flors V, Heil M, Mauch-Mani B, Pieterse CM, Pozo MJ, Ton J, van Dam NM, Conrath U (2016) Recognizing plant defense priming. Trends in plant science **21:** 818-822

Matzke MA, Mosher RA (2014) RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nature Reviews Genetics **15:** 394-408

Mauch-Mani B, Baccelli I, Luna E, Flors V (2017) Defense priming: an adaptive part of induced resistance. Annual Review of Plant Biology **68:** 485-512

Meksem K, Pantazopoulos P, Njiti V, Hyten L, Arelli P, Lightfoot D (2001) 'Forrest'resistance to the soybean cyst nematode is bigenic: saturation mapping of the Rhg1and Rhg4 loci. TAG Theoretical and Applied Genetics **103:** 710-717

Mitchum MG (2016) Soybean resistance to the soybean cyst nematode *Heterodera glycines*: an update. Phytopathology **106:** 1444-1450

Miura A, Nakamura M, Inagaki S, Kobayashi A, Saze H, Kakutani T (2009) An Arabidopsis jmjC domain protein protects transcribed genes from DNA methylation at CHG sites. The EMBO journal **28:** 1078-1086

Mosher RA, Melnyk CW, Kelly KA, Dunn RM, Studholme DJ, Baulcombe DC (2009) Uniparental expression of PolIV-dependent siRNAs in developing endosperm of Arabidopsis. Nature **460:** 283-286

Niederhuth CE, Schmitz RJ (2017) Putting DNA methylation in context: from genomes to gene expression in plants. Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms **1860:** 149-156

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

Rambani A, Rice JH, Liu J, Lane T, Ranjan P, Mazarei M, Pantalone V, Stewart CN, Staton M, Hewezi T (2015) The methylome of soybean roots during the compatible interaction with the soybean cyst nematode. Plant physiology **168:** 1364-1377

Regulski M, Lu Z, Kendall J, Donoghue MT, Reinders J, Llaca V, Deschamps S, Smith A, Levy D, McCombie WR (2013) The maize methylome influences

mRNA splice sites and reveals widespread paramutation-like switches guided by small RNA. Genome research **23:** 1651-1662

Reinders J, Wulff BB, Mirouze M, Marí-Ordóñez A, Dapp M, Rozhon W, Bucher E, Theiler G, Paszkowski J (2009) Compromised stability of DNA methylation and transposon immobilization in mosaic Arabidopsis epigenomes. Genes & Development **23:** 939-950

Rigal M, Becker C, Pélissier T, Pogorelcnik R, Devos J, Ikeda Y, Weigel D, Mathieu O (2016) Epigenome confrontation triggers immediate reprogramming of DNA methylation and transposon silencing in *Arabidopsis thaliana* F1 epihybrids. Proceedings of the National Academy of Sciences **113:** E2083-E2092

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26:** 139-140

Saze H, Kakutani T (2007) Heritable epigenetic mutation of a transposon-flanked Arabidopsis gene due to lack of the chromatin-remodeling factor DDM1. The EMBO Journal **26:** 3641-3652

Saze H, Shiraishi A, Miura A, Kakutani T (2008) Control of genic DNA methylation by a jmjC domain-containing protein in *Arabidopsis thaliana*. science **319:** 462-465

Saze H, Tsugane K, Kanno T, Nishimura T (2012) DNA methylation in plants: relationship to small RNAs and histone modifications, and functions in transposon inactivation. Plant and Cell Physiology **53:** 766-784

Schmitz RJ, He Y, Valdés-López O, Khan SM, Joshi T, Urich MA, Nery JR, Diers B, Xu D, Stacey G (2013) Epigenome-wide inheritance of cytosine methylation variants in a recombinant inbred population. Genome research **23:** 1663-1674

Schmitz RJ, Schultz MD, Lewsey MG, O'Malley RC, Urich MA, Libiger O, Schork NJ, Ecker JR (2011) Transgenerational epigenetic instability is a source of novel methylation variants. Science **334:** 369-373

Slotkin RK, Martienssen R (2007) Transposable elements and the epigenetic regulation of the genome. Nature Reviews Genetics **8:** 272-285

Slotkin RK, Vaughn M, Borges F, Tanurdžić M, Becker JD, Feijó JA, Martienssen RA (2009) Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell **136:** 461-472

Smit A, Hubley R, Green P (2015) RepeatMasker Open-4.0. 2013–2015. Institute for Systems Biology. http://repeatmasker/. org

Smith ZD, Meissner A (2013) DNA methylation: roles in mammalian development. Nature Reviews Genetics **14:** 204-220

Springer NM, Schmitz RJ (2017) Exploiting induced and natural epigenetic variation for crop improvement. Nature Reviews Genetics ***18*.9:** 563-575.

Sun Z, Guo T, Liu Y, Liu Q, Fang Y (2015) The roles of Arabidopsis CDF2 in transcriptional and posttranscriptional regulation of primary microRNAs. PLoS genetics **11:** e1005598

Supek F, Bošnjak M, Škunca N, Šmuc T (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. PloS one **6:** e21800

Taudt A, Colomé-Tatché M, Johannes F (2016) Genetic sources of population epigenomic variation. Nature Reviews Genetics **17:** 319-332

To TK, Saze H, Kakutani T (2015) Focus on Chromatin/Epigenetics: DNA Methylation within Transcribed Regions. Plant physiology **168:** 1219

Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics **25:** 1105-1111

Wang H, Beyene G, Zhai J, Feng S, Fahlgren N, Taylor NJ, Bart R, Carrington JC, Jacobsen SE, Ausin I (2015) CG gene body DNA methylation changes and evolution of duplicated genes in cassava. Proceedings of the National Academy of Sciences **112:** 13729-13734

Wang X, Hu L, Wang X, Li N, Xu C, Gong L, Liu B (2016) DNA methylation affects gene alternative splicing in plants: an example from rice. Molecular plant **9:** 305-307

Wibowo A, Becker C, Marconi G, Durr J, Price J, Hagmann J, Papareddy R, Putra H, Kageyama J, Becker J (2016) Hyperosmotic stress memory in Arabidopsis is mediated by distinct epigenetically labile sites in the genome and is restricted in the male germline by DNA glycosylase activity. Elife **5:** e13546

Zilberman D, Coleman-Derr D, Ballinger T, Henikoff S (2008) Histone H2A. Z and DNA methylation are mutually antagonistic chromatin marks. Nature **456:** 125-129

Appendix

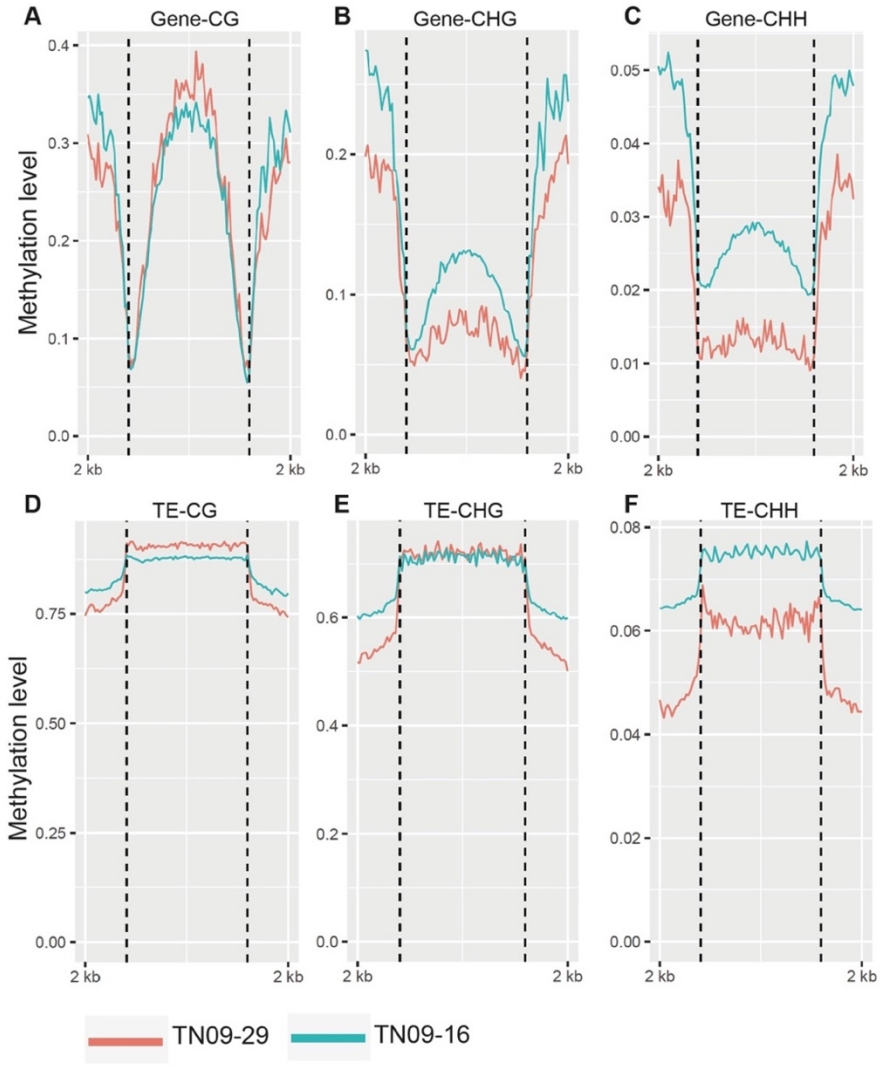

Figure 3.1 Comparison of global DNA methylation levels between TN09-16 and TN09-29 over genes and TEs in various sequence contexts under non-infected conditions. Global DNA methylation levels over protein-coding genes (A-C) and TEs (D-F) in the CG, CHG, and CHH sequence contexts. Global methylation is calculated at base pair level by dividing number of cytosines by total coverage. Promoter region of 2 kb is divided into 60 bins and methylation is averaged across each bin to be plotted as a data point on a line graph. The x-axis is methylation level represented within range from 1-0 and y-axis is the region over which methylation level was calculated.

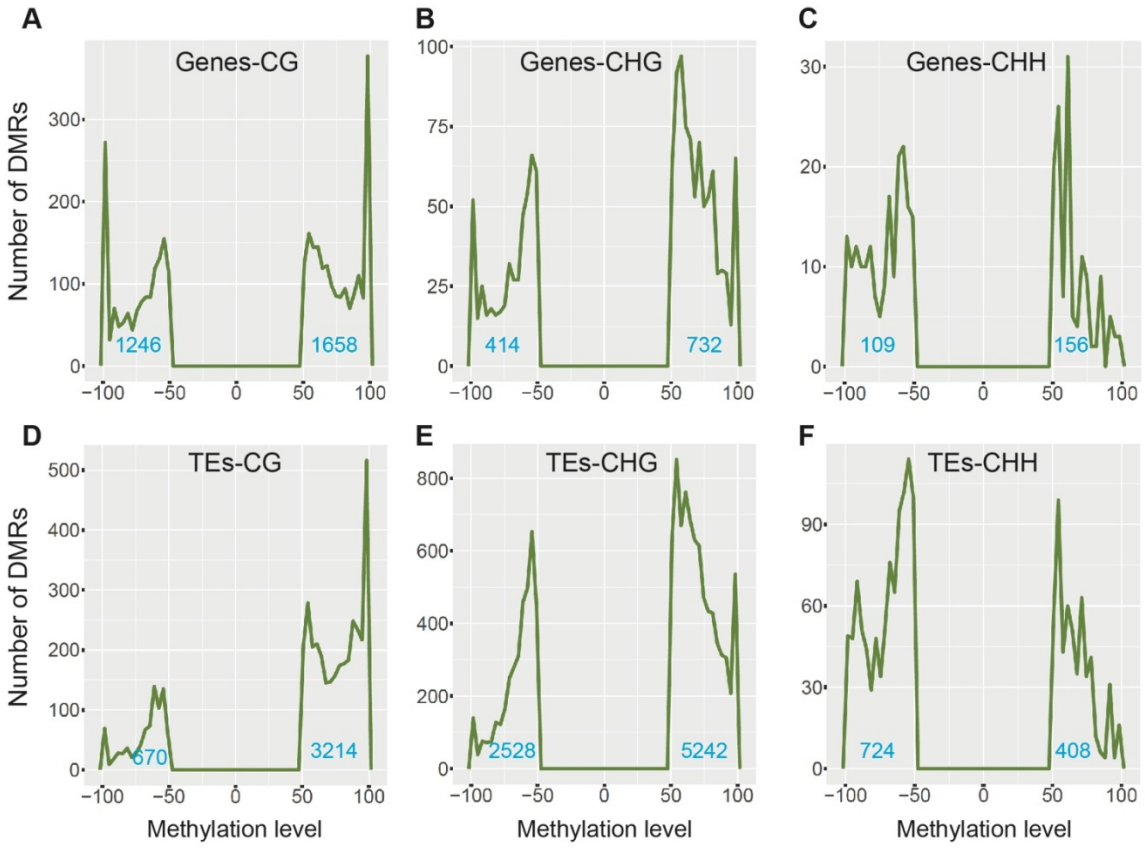

Figure 3.2 Characterization of the differentially methylated regions between TN09-16 and TN09-29. A-C, Differentially methylated regions overlapping with protein-coding genes in the CG, CHG, and CHH sequence contexts. D-F, Differentially methylated regions overlapping with TEs in the CG, CHG, and CHH sequence contexts.

Figure 3.3 Functional classification of the differentially methylated and differentially expressed genes between TN09-16 and TN09-29. A and B: Gene Ontology enrichment analysis of the differentially methylated genes (A) and differentially expressed genes (B). C: Venn diagram demonstrating the overlap between differentially methylated genes and differentially expressed genes. D: Genome browser image showing DNA hypermethylation in the promoter and gene body regions of Glyma.09G133900 in TN09-016. The hypermethylation of Glyma.09G133900 was associated with gene downregulation. Promoter region, 1 kb upstream of ATG, is highlighted.

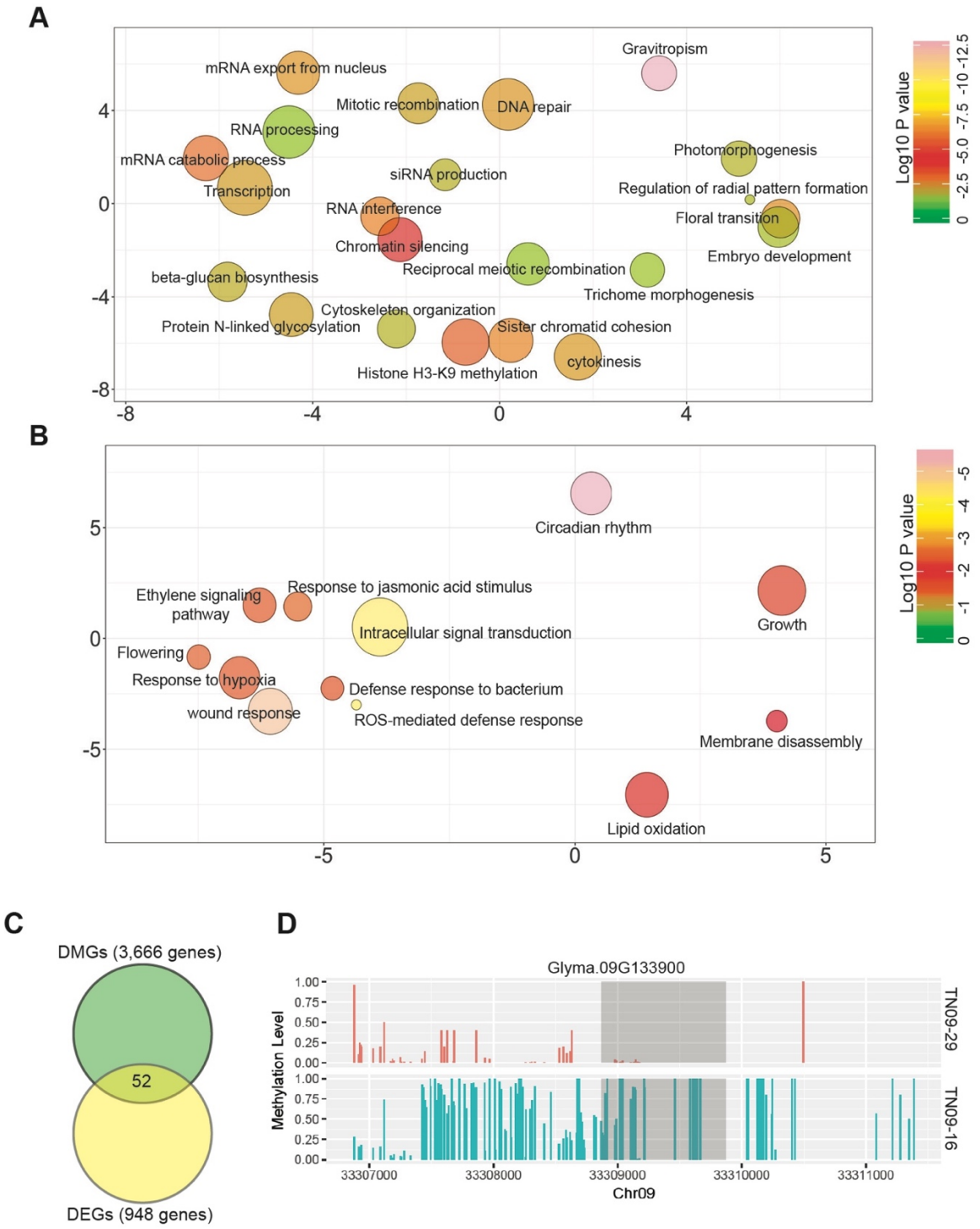

Figure 3.3 continued

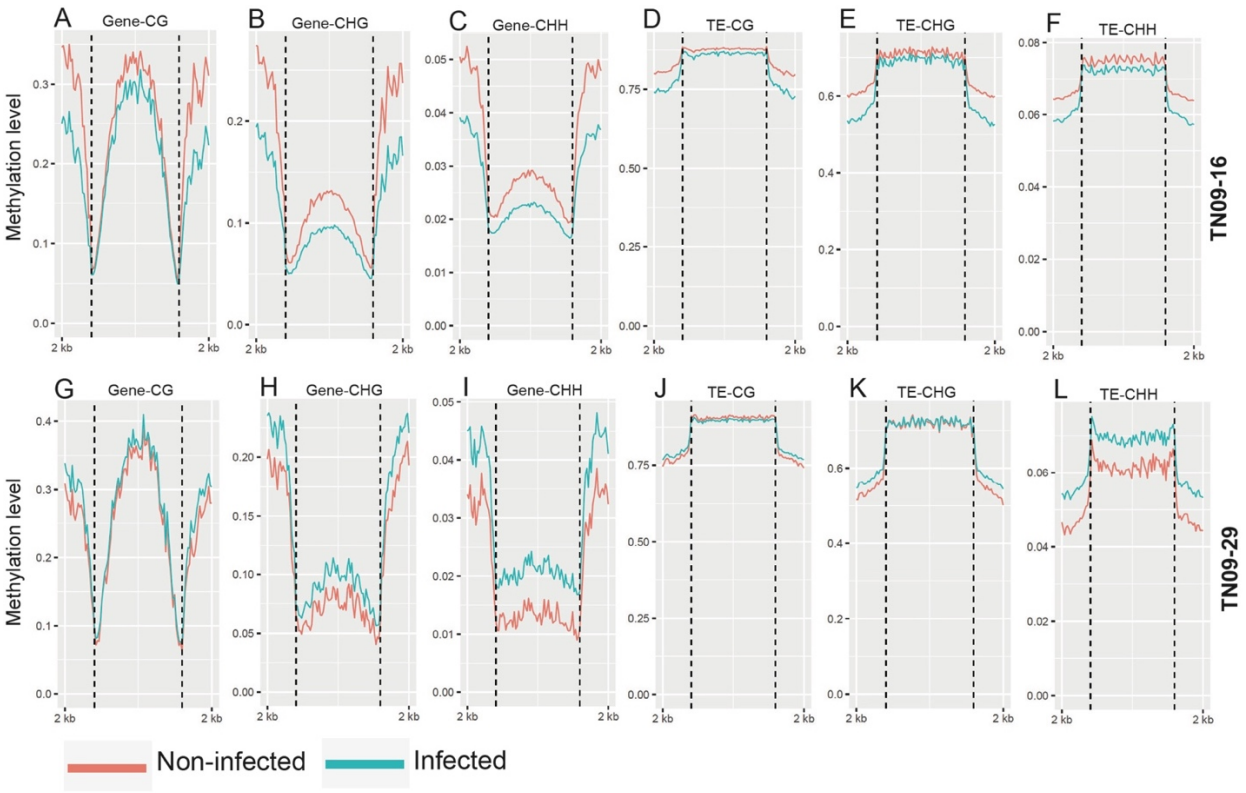

Figure 3.4 Comparison of global DNA methylation levels between the isogenic lines at 5 day post SCN infection. A-F: Global DNA methylation levels over protein-coding genes (A-C) and TEs (D-F) in infected and non-infected root samples of TN09-16. G-L: Global DNA methylation levels over protein-coding genes (G-I) and TEs (J-L) in infected and non-infected root samples of TN09-29. Global methylation is calculated at base pair level by dividing number of cytosines by total coverage. Promoter region of 2 kb is divided into 60 bins and methylation is averaged across each bin to be plotted as a data point on a line graph. The x-axis is methylation level represented within range from 1-0 and y-axis is region over which methylation level was calculated.

Figure 3.5 Specificity and magnitude of DNA methylation changes induced by SCN in the isogenic lines. A and **B:** Venn diagrams showing the numbers of DMRs overlapping with protein-coding genes and TEs in TN09-16 (A) and TN09-29 (B).**C:** Venn diagram showing 74 DMRs overlapping between those identified in TN09-16 and TN09-29. **D** and **E**: Gene Ontology enrichment analysis of the DMRs-associated genes identified in TN09-16 (D) and TN09-29 (E). **F** to **K:** Methylation level, direction, sequence contexts, and numbers of the DMRs overlapping with protein-coding genes (F-H) and TEs (I-K) identified in TN09-16 (blue lines) and TN09-29 (red lines) in response to infection by soybean cyst nematode.

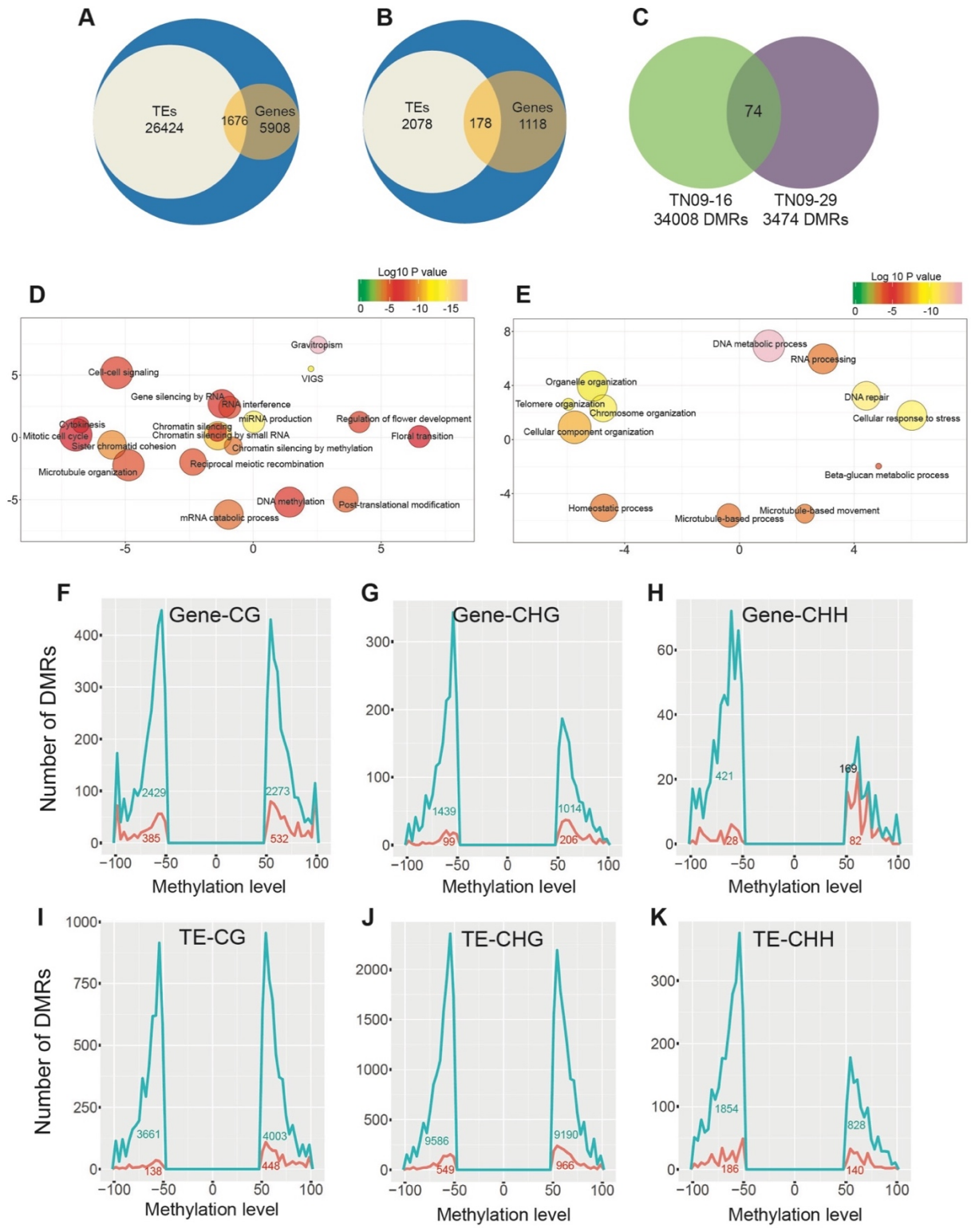

Figure 3.5 continued

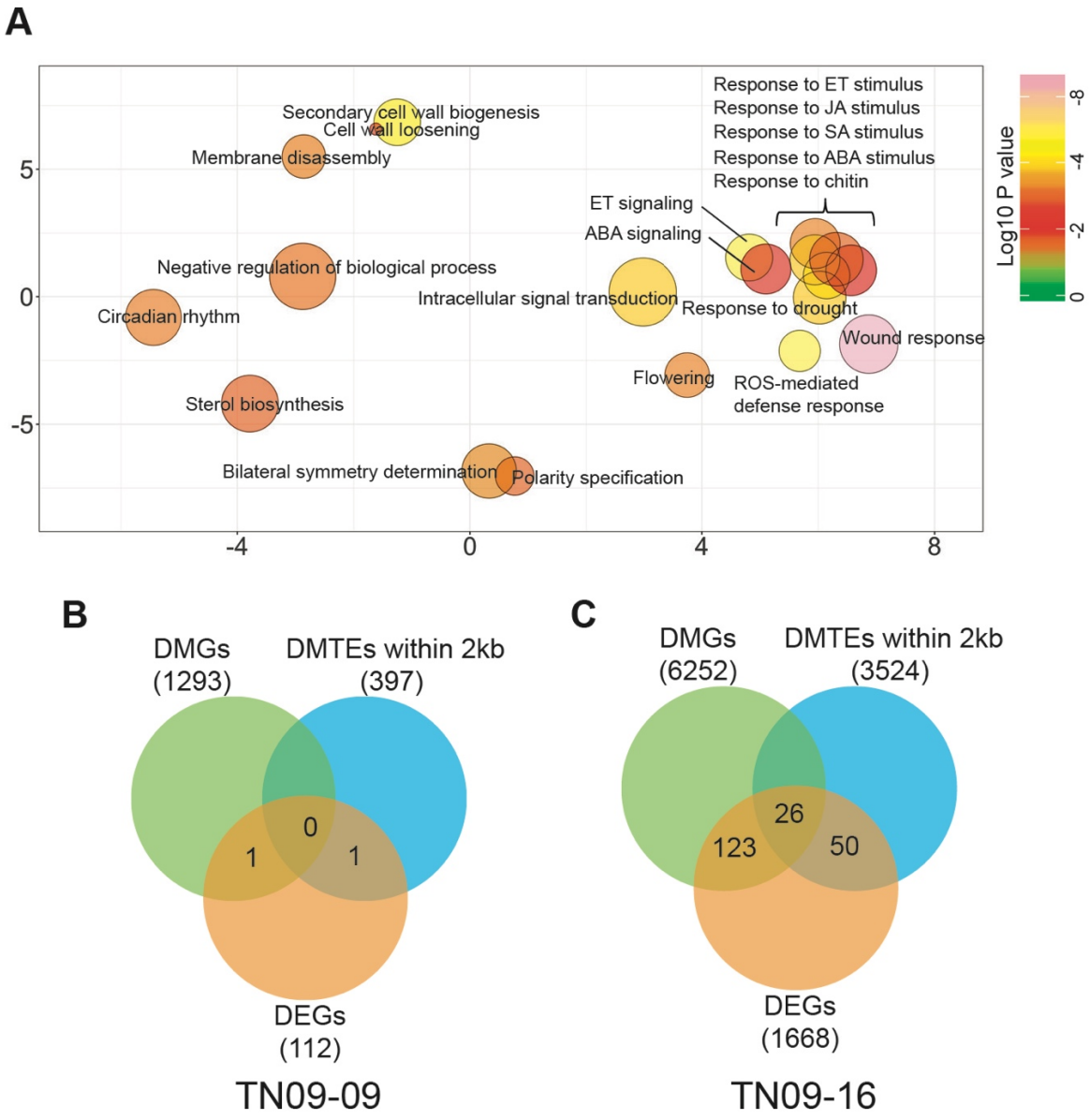

Figure 3.6 Association between DNA methylation reprograming during the susceptible interaction and gene expression changes. A: Gene Ontology enrichment analysis of the differentially expressed genes identified in the TN09-16 in response to SCN infection. **B and C:** Venn diagrams showing the overlaps between the DMGs, DMTE-associated genes, and DEGs identified in TN09-29 (B) and TN09-16 (C).

Figure 3.7 Characterization of differential methylation patterns identified in the isogenic lines that were stably inherited from the parents. A: Heat map of 59 DMRs showing the methylation patterns in TN09-29 and TN09-16 that were inherited from the parental lines Fowler and Anand, respectively. **B:** Example of the DMEs showing hypermethylation patterns in the susceptible line TN09-16 that were inherited from the susceptible parent Anand. In contrast, the resistant line TN09-29 and the resistant parent Fowler showed hypomethylation in these regions. **C and D:** Proportions for methylation contexts (C) and the overlapping genic features of the 59 DMGs (D).

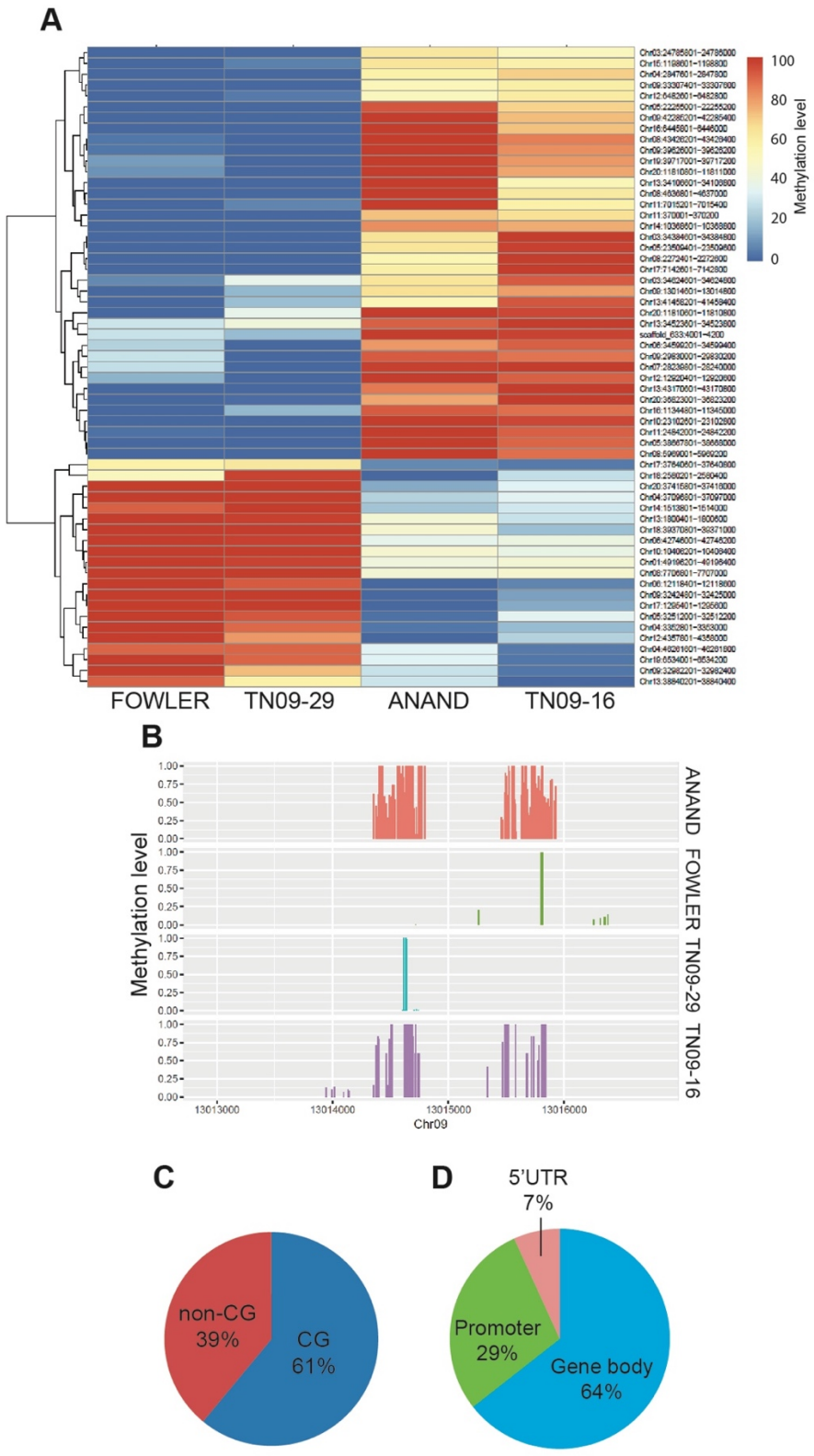

Figure 3.7 continued

Figure 3.8 Identification of novel non-parental DMRs specific to the susceptible isogenic line. A: Heat map of 56 DMRs showing methylation patterns specific to TN09-16. **B:** Example of a DMR overlapping with a protein coding gene (Glyma.08g157400) and showing methylation patterns specific to TN09-16. The region was hypermethylated in TN09-16 but hypomethylated in TN09-29 and the parental lines Fowler and Anand. **C and D:** Proportions for methylation contexts (C) and overlapping genic features of the identified 56 DMGs (D). **E:** Identification of 13 significantly differentially used exons/junctions in the RNA-seq data of TN09-16 versus TN09-29 under non-infected conditions. The mean normalized coverage of each gene was plotted against fold change values and 13 significantly differentially used exons/junctions (red dots) were identified using q value less than 0.1. **F:** Exon/junction expression profile for the Glyma.17G149600 gene. The numbers of normalized sequencing read aligned to each exon or splice junction in TN09-16 (blue) and TN09-29 (red) were obtained from RNA-seq data under non-infected condition and were displayed as gene profile plot. A gene diagram showing the location of each exon (boxes) and the predicted junction sites (dashed lines) is included below the plot. One statistically significantly (q value = 0.0072) used exon is highlighted in pink.

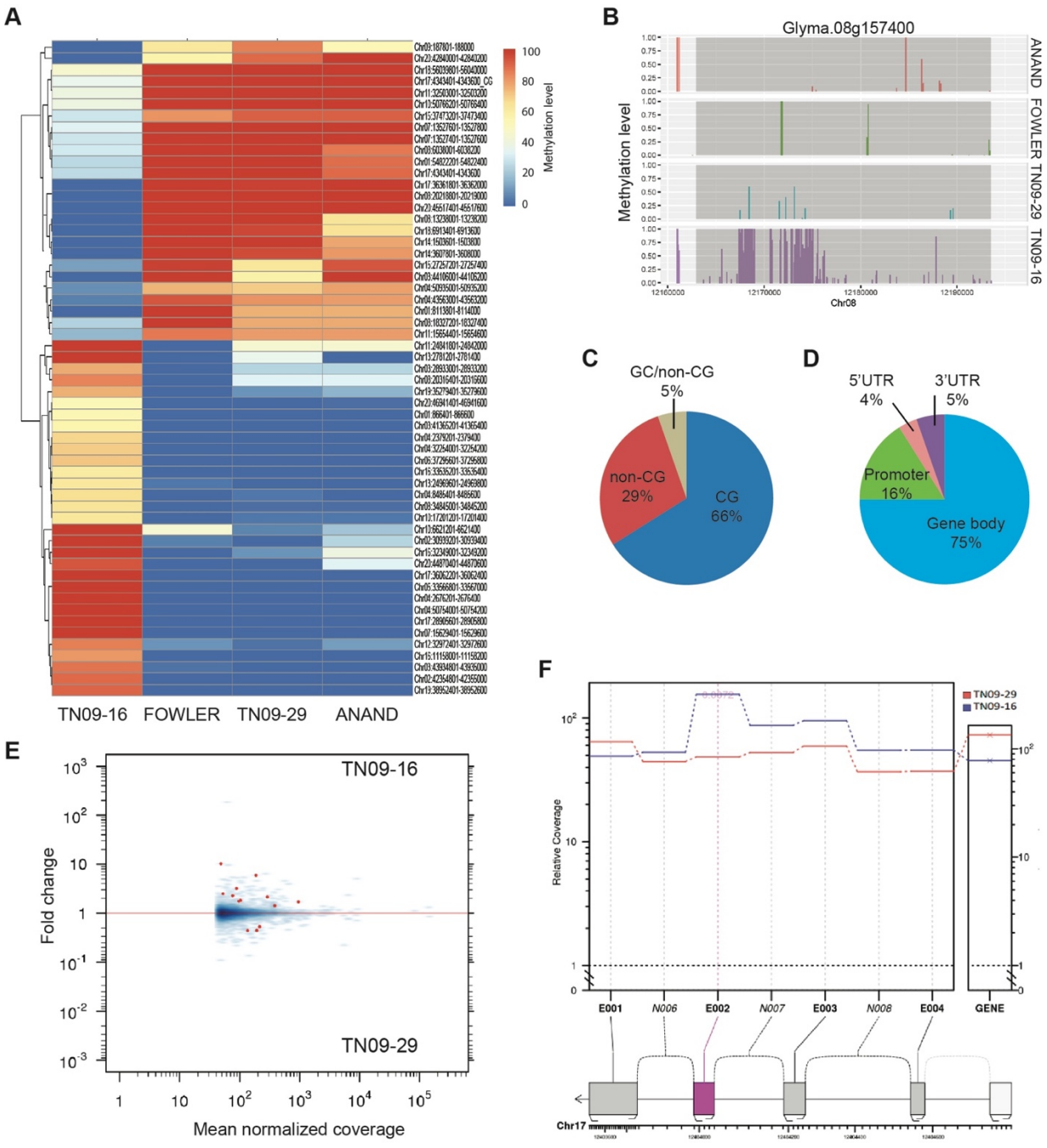

Figure 3.8 continued

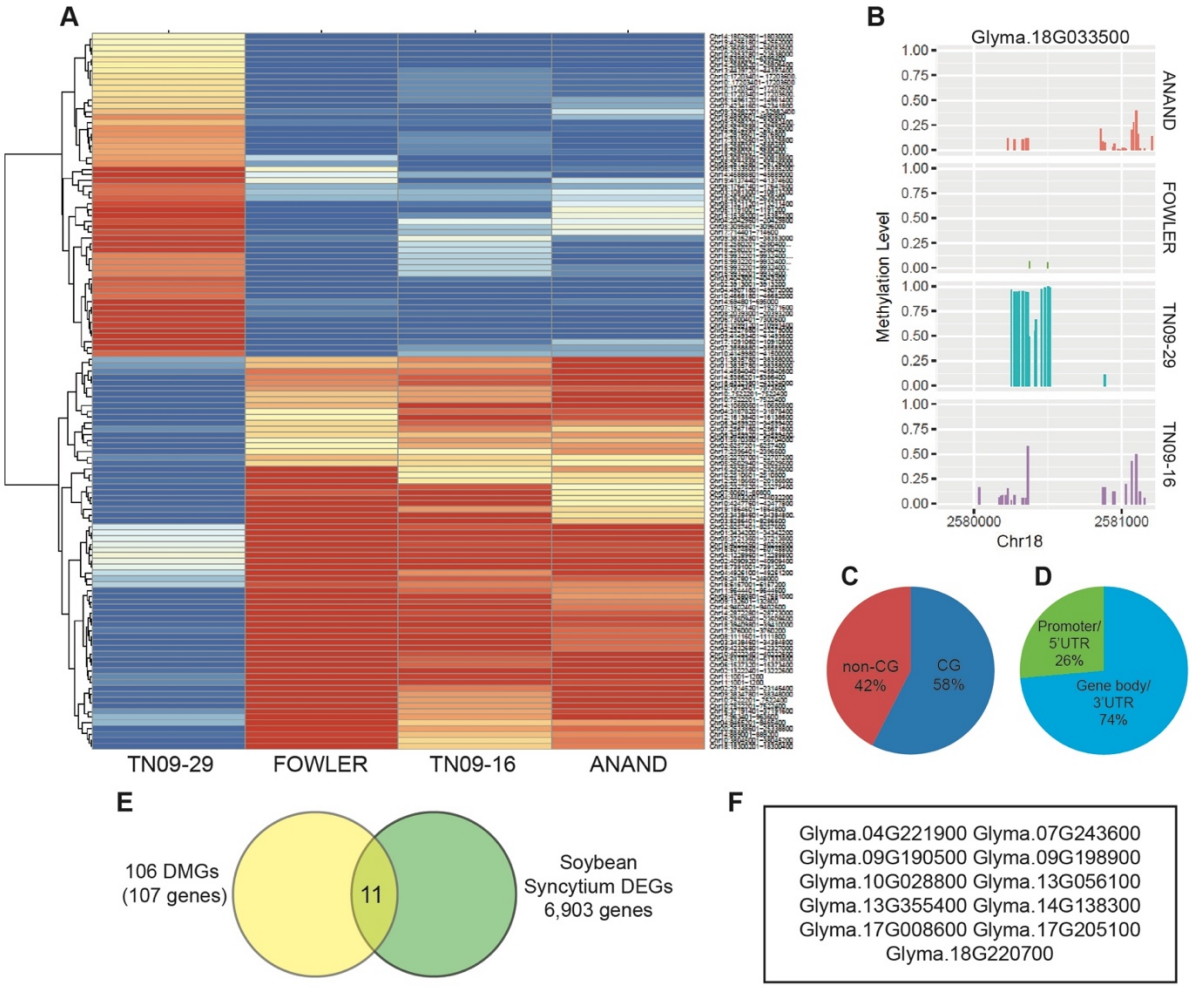

Figure 3.9 Identification of novel non-parental DMRs specific to the resistant isogenic line. A: Heat map of 106 DMRs showing methylation patterns specific to TN09-29. **B:** Example of a DMR overlapping with a protein coding gene (Glyma.18g33500) and showing methylation patterns specific to TN09-29. The region was hypermethylated in TN09-29 but hypomethylated in TN09-16 and the parental lines Fowler and Anand. **C and D:** Proportions for methylation contexts (C) and overlapping genic features of the identified 106 DMGs (D). **E.** Identification of 11 genes overlapping between the 107 DMR-associated genes and the previously identified syncytium DEGs. **F:** Accession numbers of the 11 overlapping genes.

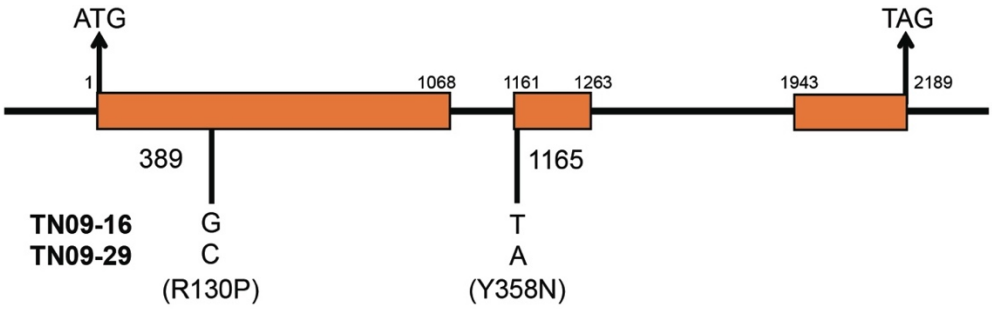

Figure S.3.1 Schematic representation of the two single nucleotide polymorphisms identified in the susceptible line TN09-16 leading to R130P and Y358N amino acid substitutions.

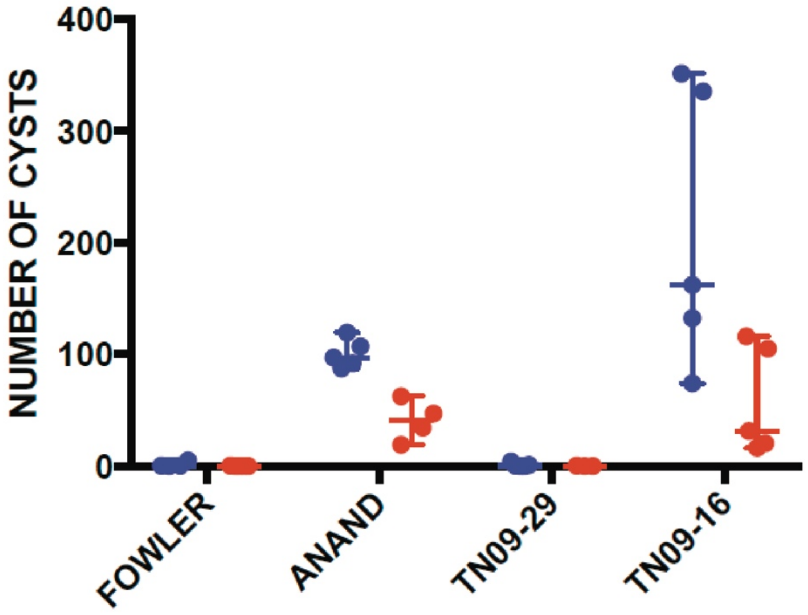

Figure S.3.2 Responses of TN09-016, TN09-29, and the parental lines Fowler and Anand to infection by the soybean cyst nematode race 3 (HG type 0).

The number of cysts was determined five weeks post inoculation. Data from two experiments (highlighted in red and blue), each with at least 10 plants per line are shown. Each data point represents the mean of cyst numbers obtained from two plants.

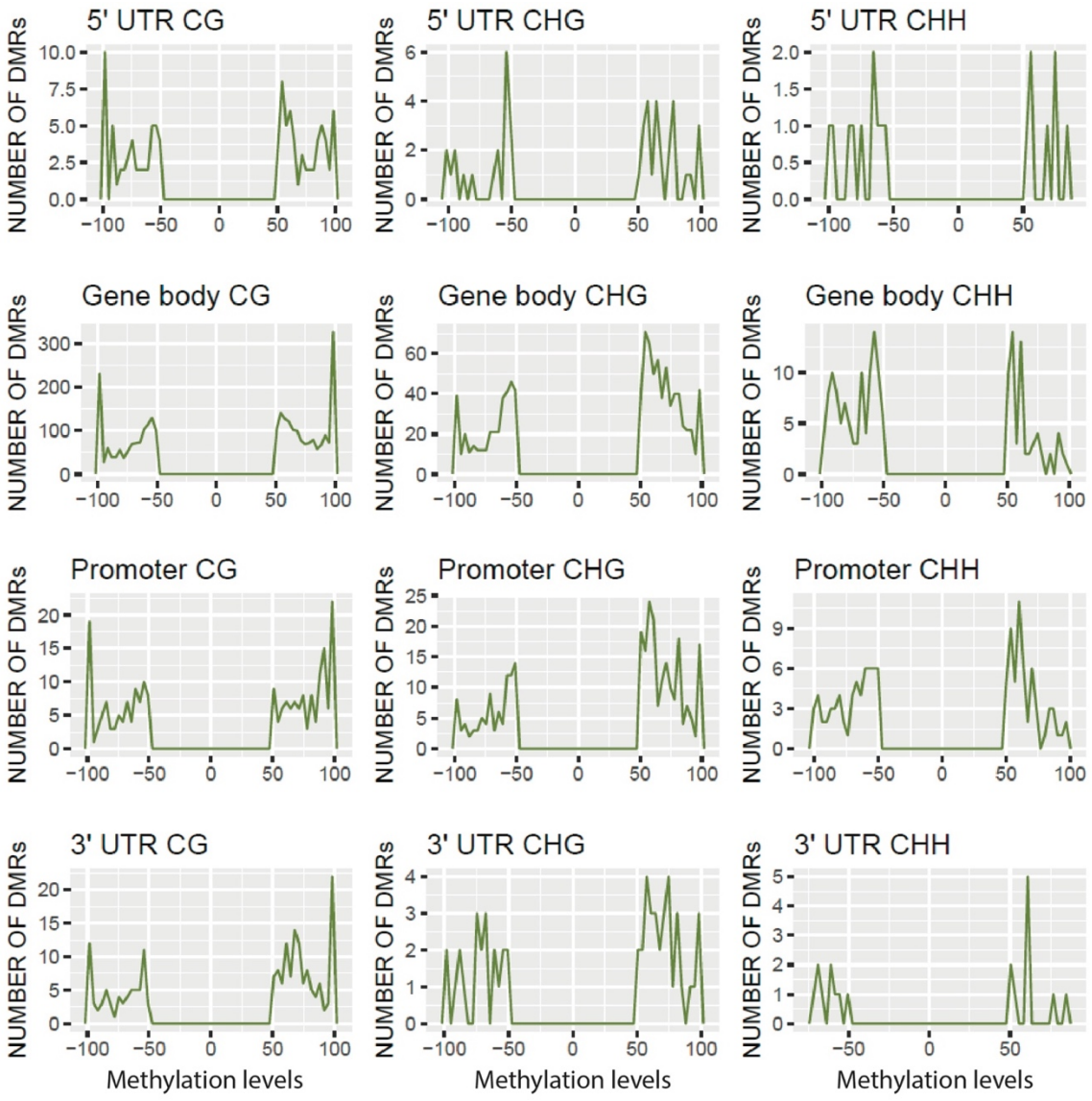

Figure S.3.3 Mapping DMRs identified between TN09-16 and TN09-29 under non-infected conditions to various annotated features of protein-coding genes, including promoter, gene body, and 5' and 3'UTRs.

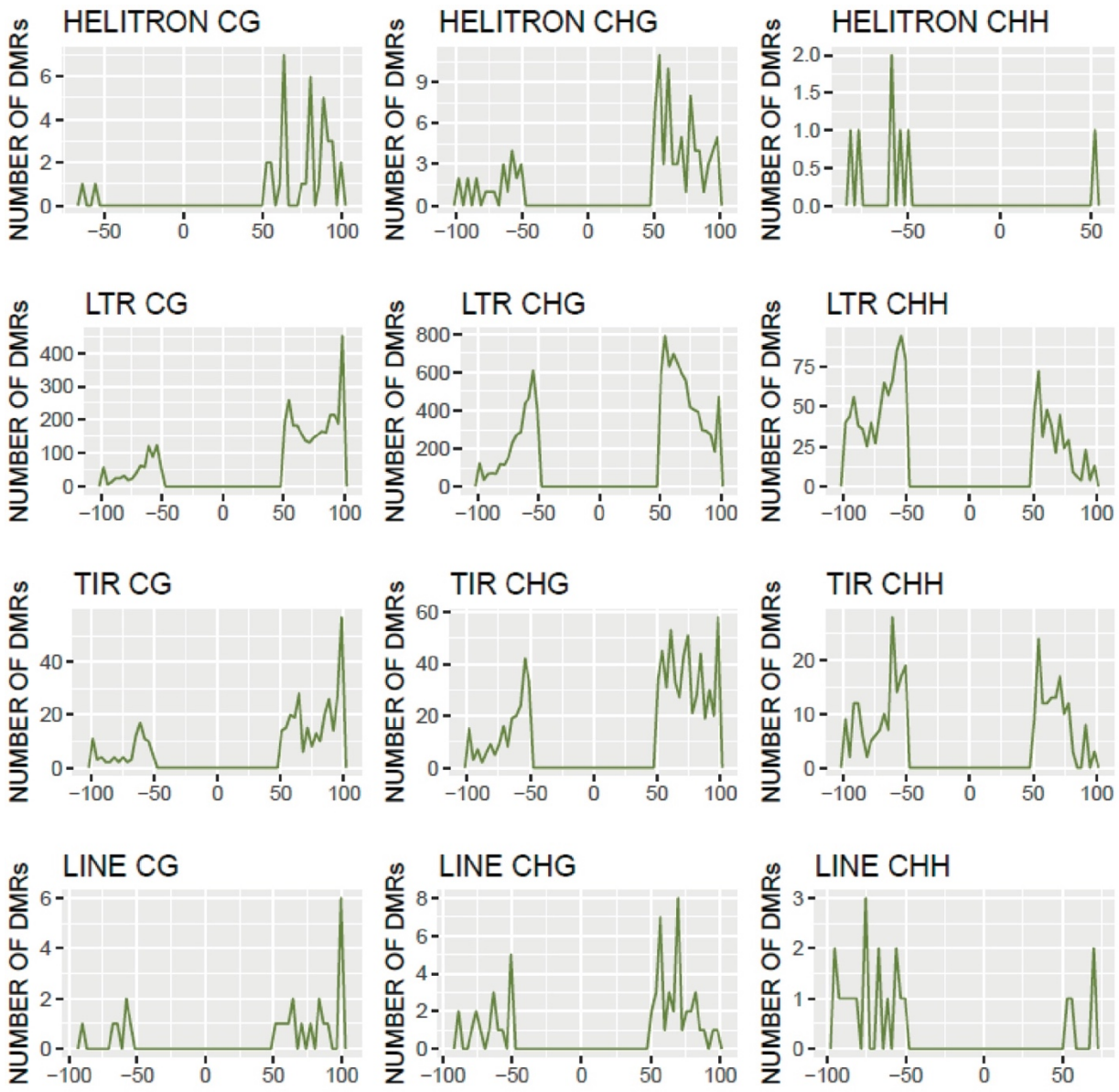

Figure S.3.4 Mapping DMRs identified between TN09-16 and TN09-29 under non-infected conditions to various transposon families, including Helitron, TIR, LTR, and Line.

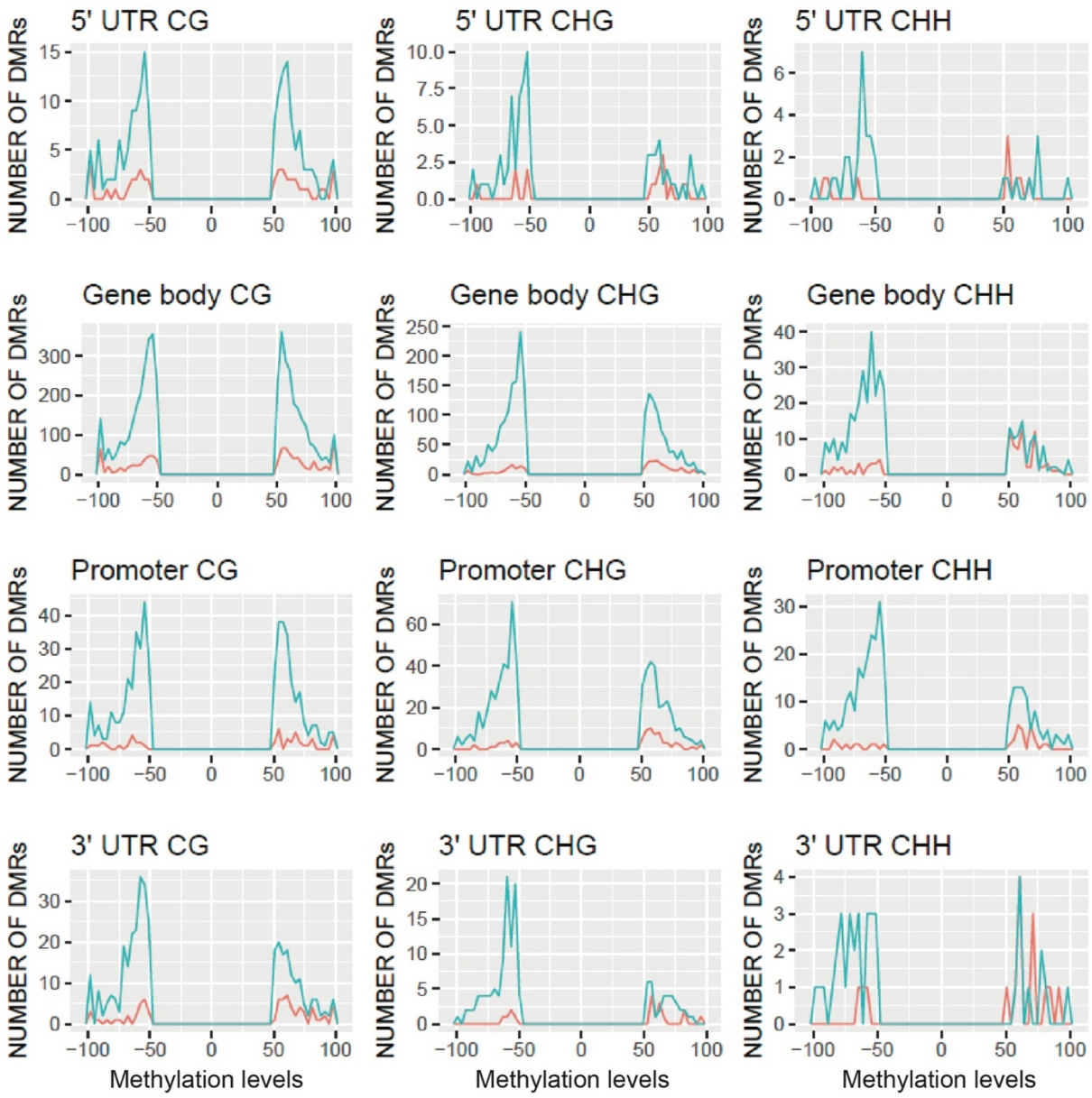

Figure S.3.5 Numbers and direction of the DMRs identified in TN09-16 (blue) and TN09-29 (red) in response to SCN infection and overlapping with protein-coding genes.

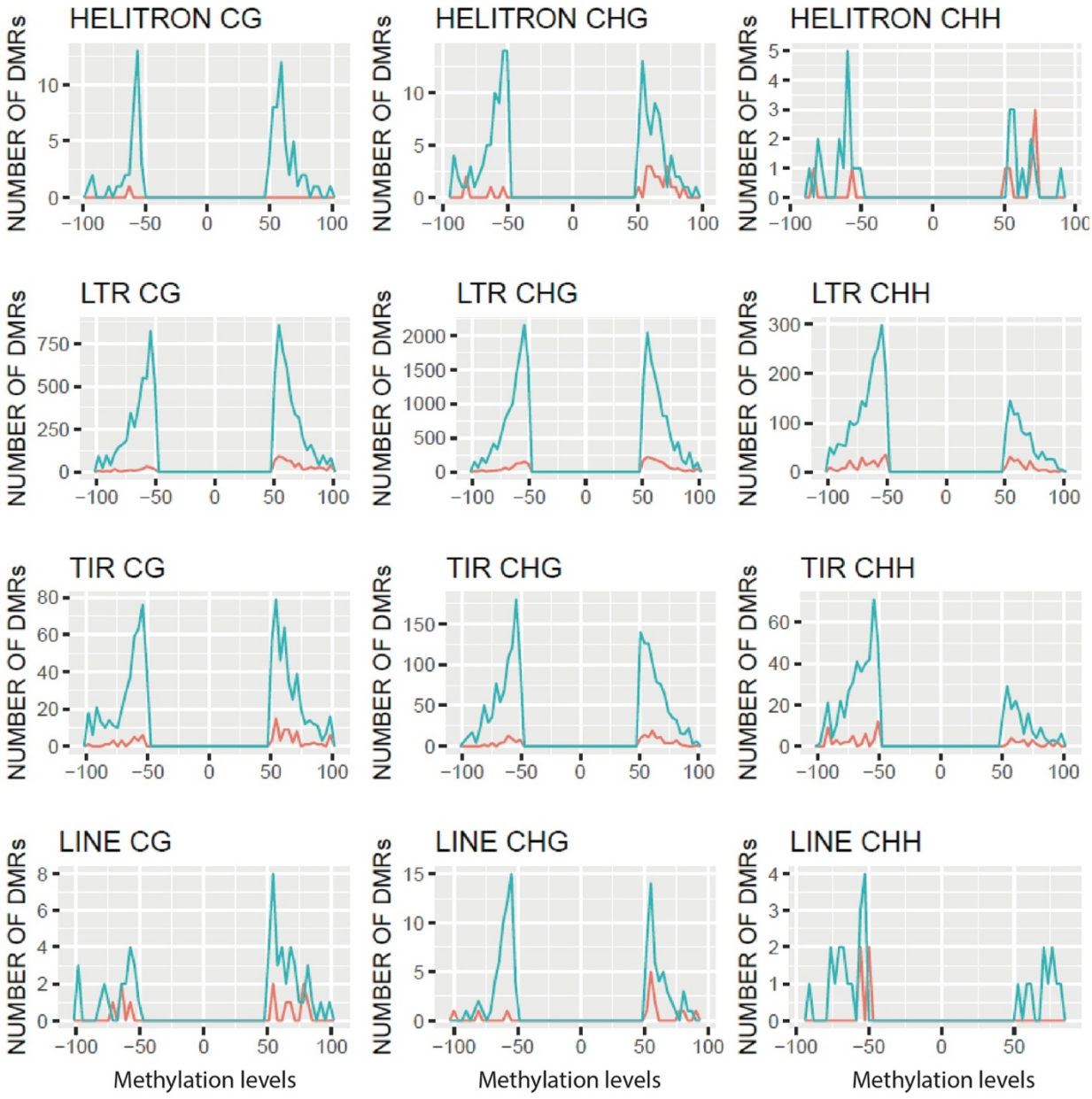

Figure S.3.6 Numbers and direction of the DMRs identified in TN09-16 (blue) and TN09-29 (red) in response to SCN infection and overlapping with various transposon families.

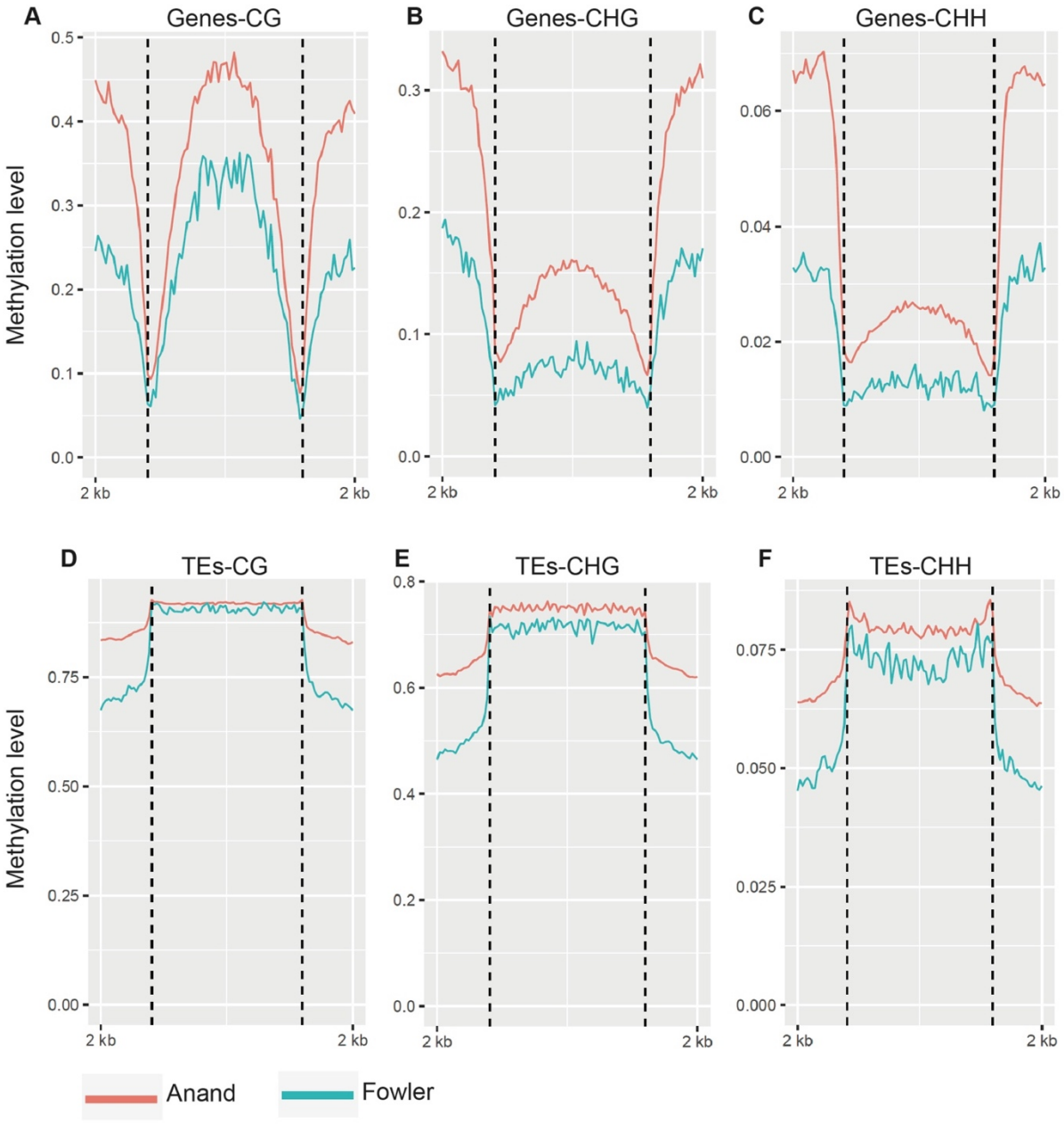

Figure S.3.7 Comparison of global DNA methylation levels between the parental lines (Anand and fowler) over protein-coding genes (A to C) and TEs (D to F) in various sequence contexts under non-infected conditions. Global methylation is calculated at base pair level by dividing number of cytosines by total coverage. Promoter region of 2 kb is divided into 60 bins and methylation is averaged across each bin to be plotted as a data point on a line graph. The x-axis is methylation level represented within range from 1-0 and y-axis is region over which methylation level was calculated.

Chapter 4: Identification of differentially methylated miRNAs during compatible and incompatible interactions between soybean and soybean cyst nematode

Running Title: Identification of differentially methylated miRNAs during compatible and incompatible interactions between soybean and soybean cyst nematode.

Aditi Rambani[1], Yangfeng Hu[1], J. Hollis Rice[1], C. Neal Stewart[1], Vince Pantalone[1] and Tarek Hewezi[1*]
[1]Department of Plant Sciences, University of Tennessee, Knoxville, Tennessee 37996

To be submitted to Plant Biotechnology Journal

I designed and performed methods to analyze methylation of miRNA genes from whole genome methylation data. I built a database from publically available soybean degradome datasets to find all known validated targets of soybean miRNAs. I conducted all the steps for validation of miRNA gene in hairy roots except cloning of the gene. I contributed towards writing the manuscript.

Abstract

microRNAs (miRNAs) are small 21-22 nt non-coding RNAs that regulate the expression of their target genes post-transcriptionally. While the function of a number miRNAs in regulating syncytium transcriptome has been recently elucidated, mechanisms that control miRNA transcription during nematode infection remain largely unknown. In this study, the potential implication of DNA methylation in controlling miRNA gene transcription was examined. DNA methylation changes in promoter and primary transcript regions of miRNA genes were quantified in susceptible and resistant lines in response to SCN infection. In the SCN-susceptible lines Williams 82 and TN09-016, 28 and 72 differentially methylated miRNA genes were identified, respectively. Using computational tools and publically available degradome datasets we identified target genes for these differentially methylated miRNAs. Although, only 8 miRNAs were similarly differentially methylated in both lines, these genetically distant lines appear to impact different miRNA genes with similar regulatory function. In the SCN-resistant TN09-29 only 14 were identified differentially methylated miRNAs under infection. Interestingly, many of the identified targets of the differentially methylated miRNAs were among those previously reported as syncytium differentially expressed genes, suggesting a regulatory function of syncytium gene expression. Functional validation of miR9746 through overexpression in hairy roots confirmed its role in facilitating SCN parasitism. This study is the first to explore the role DNA methylation in regulating miRNA transcription during SCN infection and provides new insights into the epigenetic mechanisms controlling miRNA function during SCN-soybean interactions.

1. Introduction

MicroRNAs (miRNAs) are a class of non-coding small RNAs (smRNAs) that play a crucial role in the regulation of gene expression at post-transcriptional levels. Initially, miRNAs were discovered using laborious and time consuming techniques like cloning (Llave et al., 2002; Sunkar and Zhu, 2004) but gradual advancement of sequencing and computational technologies has led to discovery of hundreds of miRNA genes in plants (Song et al., 2011; Shamimuzzaman and Vodkin, 2012; Fang et al., 2013; Hu et al., 2013; Arikit et al., 2014; Xu et al., 2014; Zhao et al., 2015; Ding et al., 2016; Xu et al., 2016). As the name suggests, miRNAs are short 21-22 nucleotide non-coding RNA, they are formed by processing of longer primary transcripts (pri-miRNAs) folded into a hairpin loop-like structures (Rogers and Chen, 2013). The miRNA genes are transcribed by RNA polymerase II (Pol II) and the primary miRNA transcript is stabilized by addition of poly A tail to the 3' end. After processing from DCL1, the 3' terminus of mature miRNA is methylated by HUA ENHANCER (HEN1) and exported to cytoplasm. In cytoplasm, the functional strand of the mature miRNA is loaded together with AGO2 protein into the RNA-induced silencing complex (RISC) to target mRNA transcripts containing perfect or near perfect complementary sequence for degradation, causing post-transcriptional silencing (Bartel, 2004).

Methylation of cytosine residue in DNA at regulatory region can alter the function of a gene without any change at genome sequence level. In plant genome, methylated cytosine occurs in three different contexts—CG, CHG, and CHH, where H denotes any base other than guanine (G), the regulatory role of cytosine methylation is context- and region-dependent (Suzuki and Bird, 2008; Feng et al., 2010). The pioneering studies in the field of epigenomics were done to understand the role of DNA methylation during cell differentiation but very soon greater importance of this epigenetic mark in genome organization and gene expression regulation was realized. Methylation at certain regions in the genome is important to establish heterochromatic regions and also for delineating centromeres, it is also a mechanism to keep transposable elements dormant (Law and Jacobsen, 2010). Such methylation marks involved in genome organization are

maintained very stably over generations. Environmental stressors are also known to induce transient non-heritable changes in DNA methylation patterns (Boyko et al., 2010; Secco et al., 2015; Yong-Villalobos et al., 2015; Wibowo et al., 2016). Stress-induced DNA methylation may be essential for a variety of reasons, but foremost, it can alter gene expression levels at a global scale, which is important to elicit a proper coping mechanism. Global gene expression change is also necessary for successful infection by phytopathogens.. Methylome changes induced by phytopathogens have been reviewed in detail (Hewezi et al., 2017); there is a common trend of global DNA hypomethylation of genome during the susceptible interactions. The hypomethylation response to pathogen infection is unique to each pathogen and impact specific genes involved in various biological processes and plant defense.

The interaction between soybean and soybean cyst nematode (SCN, *Heterodera glycines*) offers an excellent system to study the methylome changes induced in response to pathogen. Being an obligate biotroph, sedentary endoparasitic, SCN infects and establishes a feeding site, called syncytium, in the root tissues without causing necrosis or cell death. This offers more clarity in recognition of true methylation patterns associated with plant-pathogen interaction, as cell death response can muddle the methylation profile of the tissue. SCN is also a notorious pest of soybean crop, accounting for loss of a billion dollars annually in U.S and there is a dire need to discover novel mechanisms of resistance in soybean. We built a whole genome DNA methylation map of Glycine max cv. Williams82 roots undergoing compatible interaction with SCN to better understand epigenetic dynamics of the process (Rambani et al., 2015). Infected roots undergo global hypomethylation in all sequence contexts. More importantly; differential methylation is induced in various genic regions associated with genes known to be differentially expressed in the syncytium. Differential methylation in the regulatory promoter region of the genes has an impact on transcript abundance in the infected root tissues. These results indicate that along with other mechanisms, DNA methylation also plays an important regulatory role in determining soybean susceptibility to SCN. To uncover methylation patterns unique to compatible and incompatible SCN interactions, methylation profiles of resistant and susceptible near-isogenic soybean lines under SCN infection were compared (Chapter 3). In

accordance with our previous findings, global hypomethylation in all sequence contexts was observed in the susceptible line, but interestingly, opposite was observed in the resistant line. In the resistant line, a global trend of hypermethylation emerged for non-CG context and little methylation difference was observed in CG context. These findings further cement the importance of DNA methylation in regulating soybean response to SCN infection.

The efforts of previous studies exploring pathogen induced methylation changes whether in soybean or in Arabidopsis have been solely focused on genic and repeat regions (Dowen et al., 2012; Rambani et al., 2015; Hewezi et al., 2017). A much important intergenic feature of the genome, namely, the smaller genes coding for primary transcripts of miRNAs have not been investigated in this respect. Direct role of DNA methylation in regulating the expression of miRNAs in cancer cell lines has been firmly established for many different types of cancers including, for example, colorectal (Yan et al., 2011), leukemia (Hale et al., 2017), breast (Lehmann, 2014; Pronina et al., 2017), gastric (Li et al., 2015). Despite the large number of miRNA genes showing differential expression under various stress conditions (Sunkar et al., 2012), a similar role of DNA methylation in regulating expression of miRNA genes in defense response has not been reported in plants. Song et al. recently reported a negative correlation between DNA methylation of 5' flanking region and miRNA expression during bisexual flower development in popular (Song et al., 2015). This led to the discovery of key miRNA genes regulated by DNA methylation during flower development. In this chapter, we report the first analysis of DNA methylation changes induced specifically in the promoter of miRNA genes during SCN infection. The possible mechanistic aspects of SCN resistance conferred by targets of differentially methylated miRNAs were also investigated. We discovered many potential candidate miRNA genes that are involved in modulating soybean susceptibility through our differential methylation analysis and validated the role of one miRNA.

2. Results

2.1 Identification of differentially methylated miRNA genes in Williams 82 during SCN infection

Soybean epigenome changes significantly during compatible interaction with SCN and imparts another layer of gene regulation for mechanisms controlling soybean response to infection (Rambani et al., 2015). To further dissect the basis of this genetic regulation, we used our previously generated Williams 82 whole genome bisulfite sequence (BS) data to investigate methylation changes occurring specifically in microRNA genes during the susceptible interaction with SCN. The genomic coordinates of all known 597 soybean miRNAs were downloaded from miRBase (Release21; http://www.mirbase.org/). MiRBase coordinates are annotated in *Glycine max* assembly version 1 (http://www. phytozome.net/), therefore, all the BS reads were mapped to the corresponding genome assembly to identify differential methylation. Promoter was defined as 2 kb upstream of the primary miRNA sequence because miRNA promoters are known to fall between that range (Monteys et al., 2010) .We plotted global DNA methylation levels under control and infected conditions at 5 days post infection (dpi), across the promoter region of all miRNA genes. A trend of global DNA hypomethylation was observed in the infected root samples compared to non-infected controls in all methylation sequence contexts (Figure 4.1 A,-C).

To identify differentially methylated miRNA promoter, the 2 kb promoter region was split into smaller tiles of 500 bp and compared between control and SCN-infected samples. Using this approach, we identified 28 miRNAs that are significantly (adjusted p-value < 0.01) differentially methylated (>25% methylation difference) in the promoters during SCN infection (Figure 4.1b). miRNA gene promoters were differentially methylated in the CG (46%),CHG (37%) and CHH (17%) contexts. In all DNA methylation contexts we identified more hypomethylated miRNA promoters than hypermethylated. In contrast to protein-coding genes, differential methylation in the primary transcripts of miRNA genes was negligible. Only three miRNAs were found to

be significantly differentially methylated in their primary transcripts. Two primary transcripts were hypermethylated in the CHH or CHG context and one was hypomethylated in the CHG context. Interestingly, none of the miRNA primary transcripts was detected as differentially methylated in the CG context. For the 31 differentially methylated miRNAs, a total of 140 unique target genes were identified by computational prediction and searching publically available degradome datasets (Song et al., 2011; Shamimuzzaman and Vodkin, 2012; Fang et al., 2013; Hu et al., 2013; Arikit et al., 2014; Xu et al., 2014; Zhao et al., 2015; Ding et al., 2016; Xu et al., 2016). We computationally predicted 82 target genes for 21 miRNAs using high stringency parameters. Targets were predicted using plant small RNA analysis server PSRNATarget, with only one basepair mismatch allowed in the miRNA binding region (Dai and Zhao, 2011). We could also find 71 validated targets for 19 miRNAs in the degradome datasets. Only 7 differentially methylated miRNAs do not have any degradome validated targets in our study and for 5 miRNAs no targets were detected either computationally or in the degradome datasets. Gene Ontology (GO) classification and enrichment analysis of all identified target genes was performed using soybase database. GO terms associated with glucosinate biosynthesis, sugar catabolism, and transporter proteins, particularly for sulphates and auxin biosynthesis, were significantly enriched among the 140 target genes (Figure 4.2A and B). There was an overlap of 23 target genes with differentially-expressed genes in syncytia and these genes were targeted by 15 unique differentially methylated miRNAs belonging to 12 different miRNA families (Figure 4.2C and D).

2.2 Comparison of differential methylation in miRNA genes during compatible and incompatible interactions using two near-isogenic lines

Similarly, we identified differentially methylated miRNAs between the near-isogenic lines TN09-16 and TN09-29 under infected condition. We utilized our whole methylome data from the two near-isogenic lines TN09-16 and TN09-29 described in the previous chapter. These near-isogenic lines are nearly genetically identical but differ

in their response to SCN infection. TN09-16 is highly susceptible and TN09-29 is resistant to SCN race 3. All the BS reads from the two lines were mapped to the *Glycine max* assembly version 1 to correspond with MiRBase annotation coordinates. We compared global DNA methylation levels over the promoter (2 kb upstream of the primary transcript) of all the annotated miRNA genes between the isogenic lines under infected conditions (Figure 4.3 A-C). Using the same parameters described above, we discovered a significant number of differentially methylated miRNAs particularly in the susceptible line TN09-16. While only 15 significantly differentially methylated miRNAs were identified in the resistant line in response to SCN infection, 72 differentially methylated miRNAs were identified in the susceptible line. Differential methylation of the 15 miRNAs identified in the resistant line was found in CG (22%), CHG (50%), and CHH (28%) with 12 miRNAs hypermethylated and 6 hypomethylated. The differentially methylated miRNA promoters identified in the susceptible lines were found in CG (27%), CHG (49%), and CHH (24%), with 46 miRNAs being hypermethylated and 67 hypomethylated (Figure 4.4 A).

More miRNA genes were hypomethylated than hypermethylated in susceptible line during SCN infection, contrary to resistant line (Figure 4.4 A). This trend arose mainly due to opposite direction of differential methylation in CHG context. In both lines, ~50% of differentially methylated microRNA genes were in the CHG context and the majority of these were hypermethylated in the resistant line. Margin of difference between hyper- and hypomethylation in miRNA genes was much less for CG context in both lines. Both lines showed higher number of hypomethylated miRNAs in the CHH context compared to CG and CHG contexts (Figure 4.4 A). Most of the methylation change occurred in the promoter region for the resistant line; only 1 miRNA out of the 15 was found to be significantly differentially methylated in gene body. In the susceptible line, however, 26 miRNAs were differentially methylated in gene body, 53 in promoter, and 7 both in gene body and promoter regions. There were five differentially-methylated miRNAs that were common to both lists for resistant and susceptible lines (Figure 4.4 B). Interestingly, 2 out of the 5 common miRNAs had methylation change in the same direction in both lines, whereas the remaining 3 showed methylation change in the opposite direction (Figure 4.4 C, D).

2.3 Identification and functional classification of differentially methylated miRNA-targeted genes for compatible and incompatible interactions

Potential targets of differentially methylated microRNA during compatible and incompatible SCN reaction were predicted using the webserver, psRNATarget (Dai and Zhao, 2011). The targets were predicted with high stringency parameters, therefore loosely matched targets were dropped. We predicted 297 targets for 62 out of total 72 differentially methylated miRNA in TN09-16 susceptible line and 78 targets for 13 out of total 15 differentially methylated miRNA in TN09-29 resistant line. Using soybase GO enrichment tool we discovered 35 significantly enriched GO terms for targets of differentially methylated microRNA in TN09-16 susceptible line. Twenty-four of 35 were in biological function category broadly falling into processes like cell differentiation, sulfate transport, amine metabolic process and auxin biosynthesis (Figure 4.5 A). A total of 12 GO terms were found significantly enriched amongst the targets of differentially methylated miRNA identified in the resistant line. These GO terms are related to amine metabolic process, maintenance of apical meristem identity, root hair cell tip growth, microgametogenesis, seed development and somatic embroyogenesis (Figure 4.5 B). Amine metabolic process was the only common enriched GO category to both lines. Differential methylation of microRNA during compatible and incompatible interactions might affect expression of its target genes. To explore this aspect, we investigated whether these target genes are known to be differentially expressed in the syncytium. We found that 40 genes targeted by 18 differentially methylated miRNAs in the susceptible line are among those previously reported as syncytium differentially expressed genes (Figure 4.6 A). Many of these genes function as transcription factors containing DNA binding site (Figure 4.6C). In the resistant line we found 5 target genes targeted by 3 different miRNA families overlapping with syncytium differentially expressed genes, these target genes were also mainly transcription factors (Figure 4.6 B, D).

2.4 Differential methylation of miRNA9746 gene family

In soybean miR9746 exists in a gene family with 9 members located at different chromosomes. All miR9746 members share the same 21-nt mature sequence. In addition, 6 members share the primary transcript sequences as well (Figure 4.7 A and B). Interestingly, four of these six members were found to be differentially methylated in response to SCN infection in Williams 82. This was the only case where more than one member from the same gene family was found to be significantly differentially methylated in our study. Two members, *miR9746b* and *miR9746c,* were hypermethylated in the CG and CHG context, respectively. *miR9746e* was both hyper and hypomethylated in CG context and *miR9746f* was hypomethylated in the CHG context and hypermethylated in CG context. The differentially methylated miR9746 members seem to regulate the expression of genes involved in sugar metabolic pathways. Three predicted target genes coding for invertase enzymes, and one additional degradome validated target coding for dihydrolipoamide acetyltransferase were identified (Figure 4.7 C).

2.5 Overexpression of miRNA9746 in Williams 82 roots induces significant changes in susceptibility to SCN

The primary transcript sequence of the miR9746f was overexpressed under the control of a soybean ubiquitin promoter using transgenic hairy root system (Figure 4.7 D). The transgenic hairy roots were selected with the aid of the green fluorescent protein (GFP). Empty vector expressing only GFP was used as negative control (Figure 4.7 E). Hairy roots were induced at high frequency by both vectors and were screened based on presence of GFP signal. Roots lacking the signal were cleaved off from the plants (Figure 4.7 F&G). Three biological samples were collected from different plants and used to quantify the overexpression levels of miRA9746f in the transgenic hairy roots using qPCR. More than 5-fold increase in the expression of miRA9746f was detected in the

transgenic hairy roots compared to the GFP-control roots (Figure 4.7 H). We next determined the effect of miRA9746f overexpression on soybean susceptibility to SCN (race 3) infection. The transgenic hairy roots overexpressing miRA9746f or GFP were inoculated with about 3000 SCN eggs and 5 weeks post inoculation the number of cysts were counted and used as a measure to determine plant susceptibility. Data from three biological replicates, each with at least five plants, revealed a substantial increase of about 200% in SCN susceptibility of the transgenic hairy roots overexpressing miRNA9746f compared to control (Figure 4.7 I). This provides strong evidence that miRNA9746f levels in root tissue modulate plant susceptibility and its expression may be fine-tuned during SCN infection by gene regulation mechanisms like DNA methylation.

3. Discussion

MiRNAs are a subclass of hairpin siRNAs that are encoded from endogenous miRNA genes by RNA Pol II (Lee et al., 2004). MiRNAs genes exist in the euchromatic intergenic regions as well as within the intronic regions of annotated genes. MiRNAs provide a complex post transcriptional gene regulation mechanism and understanding the factors involved in the regulation of miRNA transcription itself will provide opportunities to control its function in various developmental and stress conditions. Our study is the first to report DNA methylation changes in soybean miRNA genes during SCN infection. We discovered that methylation changes in miRNA genes are similar to protein coding genes where more differential DNA methylated occurs during the compatible interaction compared to the incompatible interaction. We analyzed DNA methylation patterns in miRNA promoter using two SCN susceptible lines (Williams 82 and TN09-16). Surprisingly, methylation changes in microRNA gene promoters were fairly unique between these two lines. Only 8 miRNA genes were commonly differentially methylated, suggesting that soybean lines derived from different genetic backgrounds modulate DNA methylation pathway differently.

Using the two near isogenic lines, TN09-16 and TN09-29, allowed us to determine methylation differences between compatible and incompatible reaction during SCN infection with minimum genetic background effects. Differential DNA methylation in miRNA promoters differs significantly between the isogenic lines. Our analysis indicates that in response to SCN infection the susceptible line is more vulnerable to DNA methylation than the resistant line. While 72 miRNAs were identified as differentially methylated in the susceptible line, only 15 were found differentially methylated in the resistant line. Five miRNA genes were differentially methylated in both lines. Very interestingly, out of these five miRNAs three (mi5032, mi5043 and mi1520b) were methylated in opposite direction. Mi5043 targets a tetratricopetide-repeat thioredoxin gene known to be required for abiotic stress tolerance in Arabidopsis (Lakhssassi et al., 2012). mi5043 and mi1520b target genes contain DNA binding domain, potential transcription factors, suggesting a role in controlling gene transcription during SCN infection. It may be tentative to speculate that the opposite regulation of these three miRNAs between the resistant and susceptible lines may contribute towards their opposite response to nematode infection.

Generally, hypomethylation in promoter region is associated with gene upregulation, whereas hypermethylation results in gene downregulation. It has been reported in Arabidopsis and tomato that initial phases of infection are accompanied by downregulation of miRNAs and but at later stages of infection this trend is reversed (Hewezi et al., 2008; Zhao et al., 2015; Medina et al., 2017). However, in soybean no such analysis of miRNA gene expression during early and late stages of SCN infection was reported. SCN resistance is polygenic (Concibido et al., 2004), and is greatly affected by genetic background, and SCN HG type, therefore, the impact of hyper-and hypomethylation of miRNA gene expression is expected to differ from other plant-nematode pathosystems. Comparing the target genes of the differentially methylated miRNAs with the previously reported syncytium differentially expressed genes provided novel insight into a possible role of these miRNAs in regulating syncytium transcriptomes. Strikingly, we found that the majority of miRNA-targeted genes overlapping with syncytium differentially expressed genes are transcription factors (Figure 4.6 D, E). This observation holds true for both susceptible and resistant lines. It

is well established that miRNAs bring about global transcriptome change during nematode infection by targeting transcription factors (Hewezi et al., 2008; Hewezi et al., 2012; Li et al., 2012; Medina et al., 2017; Piya et al., 2017). Nearly half of the differentially methylated miRNA discovered in Williams 82, target a gene known to be significantly differentially expressed in syncytium (Figure 4.2 D). Some functional similarity was found amongst these 23 syncytium differentially expressed target genes. There were 4 different genes coding for F-box domain containing proteins and 4 different genes regulating transport of sulphur, iron and proteins. The remaining genes fall in categories of cell wall modification, DNA binding, metabolic processes and stress response. Although, different miRNAs were found to be differentially methylated in Williams 82 and TN09-16 during the susceptible interactions there was functional similarity between the target genes of differentially expressed genes miRNAs identified in Williams 82 and TN09-16. A number of DNA binding transcription factors, Leucine Rich Repeats (LRR), auxin response factors, sulfur transport genes and genes involved in sugar metabolism were represented amongst the targets of the differentially methylated miRNAs detected in both lines. Some of these genes have been reported to play crucial role in determining plant susceptibility to biotic stress (Hewezi et al., 2014; Tauzin and Giardina, 2014; Capaldi et al., 2015; Hewezi et al., 2015). Thus, it appears that the susceptible interaction in various genetic backgrounds is mediated through manipulation of similar biological processes controlled by different set of differentially methylated miRNAs.

To determine if differentially methylated miRNAs can modulate susceptibility, we selected one differentially methylated miRNA in Williams 82, miR9746, for functional validation using transgenic hairy root system. This miRNA was an ideal candidate for two reasons. Firstly, four members of the miR9746 gene family showed differential DNA methylation in both directions in various methylation contexts. Secondly, predicted and/or confirmed targets of miR9746 included three invertases, which are involved in sugar metabolism during the initiation of feeding site, and a dihydrolipoamide acetyltransferase, which found among the syncytium differentially expressed genes. We observed that over expression of miR9746 in transgenic hairy root increased plants susceptibility by about 200%. Thus, miR9746 overexpression-mediated downregulation

of invertase and dihydrolipoamide acetyltransferase dramatically enhanced SCN parasitism of soybean. This result is consistent with recent finding showing that nematode development was greatly enhanced on multiple Arabidopsis mutants of sucrose-cleaving enzymes, including sucrose synthases and invertases (Herbers et al., 1996; Hofmann et al., 2009; Bonfig et al., 2010; Hofmann et al., 2010; Cabello et al., 2013; Tauzin and Giardina, 2014). Invertase and acetyltransferase enzymes catalyze the breakdown of sucrose and other complex carbohydrate molecule into smaller molecules like glucose and fructose, which are considered the major source of carbohydrate input in the syncytium (Hofmann et al., 2009; Cabello et al., 2013). The exact mechanism through which downregulation of invertase and dihydrolipoamide acetyltransferase enhance SCN parasitism remains unknown. However, one can speculate that downregulation of these genes may activate other sucrose synthases and invertases, leading to augmented sugar pools of trehalose and 1-kestose, which may have beneficial impacts on nematode development as previously proposed (Hofmann et al., 2009; Cabello et al., 2013). Another possibility is that downregulation of invertases through overexpression of miR9746 may inhibit host defense response, leading to successful nematode parasitism, taking into consideration the role of sucrose signaling in plant immunity and stress response (Bolouri-Moghaddam et al., 2010; Bonfig et al., 2010; Tauzin and Giardina, 2014). In support with this possibility, the activity of various plant invertases was found to be associated with salicylic acid levels, pathogenesis-related (PR) gene expression, and antioxidant activity (Herbers et al., 1996; Bolouri-Moghaddam et al., 2010; Bonfig et al., 2010; Tauzin and Giardina, 2014).

In conclusion, our study is the first to investigate methylation changes in miRNA genes in response to SCN infection in two near isogenic lines and Williams 82. Our analyses provide novel insight on the epigenetic mechanisms, particularly DNA methylation, that regulate miRNAs gene expression and subsequently their target genes during SCN infection. Functional characterization of theses miRNAs, specifically those showing opposite methylation patterns between the susceptible and resistant interactions will determine the extent to which these regulatory factors shape SCN-soybean interaction.

4. Material and Methods

4.1 Plant, nematode and bacterium sources

TN09-16, TN09-25 and 'Williams 82' are the three soybean genotypes used in this project. SCN race 3 was used for inoculation for soybean SCN susceptibility assay. *Agrobacterium rhizogenes* strain K599 was used to induce transgenic soybean hairy roots in Williams 82.

4.2 Differential methylation analysis of miRNA genes

Methodology for generation of bisulfite sequence data for Williams 82 and near isogenic lines under SCN infection and non-infected condition is described in previous chapters (2 & 3). Trimmed and cleaned bisulfite sequencing reads were mapped to *Glycine max* assembly version 1 (http://www. phytozome.net/) using Bismark (Krueger and Andrews, 2011). Annotation of miRNA genes in soybean genome was downloaded from miRBase (Release21; http://www.mirbase.org/) to obtain coordinates of primary miRNA gene body in soybean genome assembly. The promoter of miRNA was defined 2 kb upstream of primary miRNA start site. R package methylKit was used to perform differential methylation analysis for miRNA primary transcripts and 2 kb promoter (Akalin et al., 2012). The 2 kb promoter region was further split into four bins of 500 bp for accurate detection of differential methylation. MiRNAs were considered significantly differentially methylated if the methylation difference was greater than 25% with adjusted p value less than 0.01.

4.3 Prediction of miRNA targets and their functional analysis

A web-based bioinformatics tool called psRNA Target Server (http://plantgrn.noble.org/psRNATarget/) was used to identify targets of differentially methylated miRNAs. Mature miRNA sequences of differentially methylated miRNAs were matched with sequences from all the annotated soybean genes to search for complementarity sequences. The alignment was performed with default parameters allowing only one mismatch in the binding region. An alignment score was calculated considering miRNA length, penalizing for G:U pairing, opening and extending a gap. Alignment score of less than 3.0 was considered a good match. The targets of miRNAs can also be predicted by sequencing degradome. Thus, we complied data from 8 publically available soybean degradome datasets and searched for targets of the differentially methylated miRNAs (Song et al., 2011; Shamimuzzaman and Vodkin, 2012; Fang et al., 2013; Hu et al., 2013; Arikit et al., 2014; Xu et al., 2014; Zhao et al., 2015; Ding et al., 2016; Xu et al., 2016). Gene ontology of target genes was determined using soybase toolshed (https://soybase.org/goslimgraphic_v2/dashboard.php) and visualized using REVIGO (Supek et al., 2011).

4.4 Cloning of miR9746 gene and construction of binary vectors for root transformation

The primary transcript of miR9746f was amplified from genomic DNA using primer pair containing BamHI and AscI restriction sites. The PCR conditions used were: 94°C for 2 min followed by 35 cycles at 94°C for 30 s, 57°C for 30 s and 72°C for 1 min 30 s and a final extension at 72°C for 10 min. The PCR product was cloned into pGMET vector (Promega) and transferred into *E. coli*. The miRNA9746 gene was spliced from pGMET vector using restriction enzymes BamHI and AscI and cloned into the binary vector pG2RNAi2 under control of Gmubi promoter. pG2RNAi2 contains the green fluorescent

protein (GFP) under the control of 35S promoter. Both empty pG2RNAi2 vector and pG2RNAi2 miR9746f were introduced into the K599 strain of A. rhizogenes for generating transgenic hairy roots

4.5 Generation of transgenic soybean hairy roots

Soybean seeds were surface sterilized with bleach and placed evenly spaced on a sheet of wet germination paper. Germinated seeds with good length (approx. 10 cm) of roots were transferred to 18 cell germination tray containing sterile autoclaved vermiculite. When the cotyledons became green but still unopened, approximately 4-5 days after transfer, the seedlings were injected with K599 strain of *A. rhizogenes* transformed with binary vectors. The K599 strain of A. *rhizogenes* containing binary vectors was cultured on LB plates containing 50 mg/mL kanamycin. Approximately, 1 ml of bacterial cell suspension in distilled water was injected in hypocotyl of soybean seedlings three times using 3-mL syringe needle. The injected seedlings were allowed to stabilize for one week and then transferred to single 10 L planting pots containing sterile vermiculite. The injected plants were grown in growth chamber under white fluorescent bulbs with cycle of 16h light/8h dark at 26°C. Several transgenic hairy roots emerged from the injection site within 4-5 weeks. Transgenic hairy roots were washed under tap water and non-transgenic soybean roots were cut off from the plants. Presence of GFP protein in transgenic hairy roots was detected using an epifluorescent microscope (Olympus stereo microscope model SZX12, Olympus America, Center Valley, PA). The filter was set at 470/40 nm excitation and QCapture 2.56 imaging software was used to take images. Soybean plants with at least three transgenic hairy roots were used SCN susceptibility assay. Additional transgenic roots were collected and used for RNA isolation and quantification of miR9746 levels.

4.6 SCN susceptibility assay

Three replicates, each containing five of control and miR9746 overexpression hairy root plants, were planted in three separate trays. Each tray was filled up with mixture of sterile sand soil in equal proportion (1:1). SCN were hatched in a hatching chamber and approximately 3000 J2s were used to inoculate each plant by adding 1mL of inoculum in the vicinity of soybean roots. 5-6 weeks post inoculation. All the plants were harvested and cysts were extracted. Number of cysts per plant were counted and used as a measure of susceptibility.

4.7 RNA isolation and qPCR quantifications

Total RNA was isolated from the soybean root tissues using NucleoSpin® miRNA (MACHEREY-NAGEL, Bethlehem, PA) and treated with DNAse I enzyme (Invitrogen) to remove DNA contamination. First strand of miRNA was synthesized and then quantified using Mir-X™ miRNA First-Strand Synthesis and SYBR® qRT-PCR kit (Clone Tech) The forward primer used in qPCR reaction was 5′-AAAGTGTTTGAATCTCAATTAGATAA-3′ and the reverse primer was mRQ 3’ primer provided with the kit. Small nuclear RNA U6 was used to normalize miR9746 expression levels. qPCR was run on a QuantStudio 6 Flex Real Time PCR System (Applied Biosystems) using the following program: 95°C for 3 min; followed by 40 cycles of 95°C for 30 s and 60°C for 30s. The dissociation curves were built by running following program: 95°C for 15 s and 50°C for 15s, followed by a slow gradient from 50°C to 95°C.

References

Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, Mason CE (2012) methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome biology **13:** R87

Arikit S, Xia R, Kakrana A, Huang K, Zhai J, Yan Z, Valdés-López O, Prince S, Musket TA, Nguyen HT (2014) An atlas of soybean small RNAs identifies phased siRNAs from hundreds of coding genes. The Plant Cell **26:** 4584-4601

Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. cell **116:** 281-297

Bolouri-Moghaddam MR, Le Roy K, Xiang L, Rolland F, Van den Ende W (2010) Sugar signalling and antioxidant network connections in plant cells. The FEBS journal **277:** 2022-2037

Bonfig KB, Gabler A, Simon UK, Luschin-Ebengreuth N, Hatz M, Berger S, Muhammad N, Zeier J, Sinha AK, Roitsch T (2010) Post-translational derepression of invertase activity in source leaves via down-regulation of invertase inhibitor expression is part of the plant defense response. Molecular Plant **3:** 1037-1048

Boyko A, Blevins T, Yao Y, Golubov A, Bilichak A, Ilnytskyy Y, Hollander J, Meins Jr F, Kovalchuk I (2010) Transgenerational adaptation of Arabidopsis to stress requires DNA methylation and the function of Dicer-like proteins. PloS one **5:** e9514

Cabello S, Lorenz C, Crespo S, Cabrera J, Ludwig R, Escobar C, Hofmann J (2013) Altered sucrose synthase and invertase expression affects the local and systemic sugar metabolism of nematode-infected *Arabidopsis thaliana* plants. Journal of experimental botany **65:** 201-212

Capaldi FR, Gratão PL, Reis AR, Lima LW, Azevedo RA (2015) Sulfur metabolism and stress defense responses in plants. Tropical plant biology **8:** 60-73

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

Dai X, Zhao PX (2011) psRNATarget: a plant small RNA target analysis server. Nucleic acids research **39:** W155-W159

Ding X, Li J, Zhang H, He T, Han S, Li Y, Yang S, Gai J (2016) Identification of miRNAs and their targets by high-throughput sequencing and degradome analysis in cytoplasmic male-sterile line NJCMS1A and its maintainer NJCMS1B of soybean. BMC genomics **17:** 24

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

Fang X, Zhao Y, Ma Q, Huang Y, Wang P, Zhang J, Nian H, Yang C (2013) Identification and comparative analysis of cadmium tolerance-associated miRNAs and their targets in two soybean genotypes. PLoS One **8:** e81471

Feng S, Cokus SJ, Zhang X, Chen P-Y, Bostick M, Goll MG, Hetzel J, Jain J, Strauss SH, Halpern ME (2010) Conservation and divergence of methylation patterning in plants and animals. Proceedings of the National Academy of Sciences **107:** 8689-8694

Ha M, Kim VN (2014) Regulation of microRNA biogenesis. Nature reviews Molecular cell biology **15:** 509-524

Hale V, Hale GA, Brown PA, Amankwah EK (2017) A Review of DNA Methylation and microRNA Expression in Recurrent Pediatric Acute Leukemia. Oncology **92:** 61-67

Herbers K, Meuwly P, Frommer WB, Metraux J-P, Sonnewald U (1996) Systemic acquired resistance mediated by the ectopic expression of invertase: possible hexose sensing in the secretory pathway. The Plant Cell **8:** 793-803

Hewezi T, Howe P, Maier TR, Baum TJ (2008) Arabidopsis small RNAs and their targets during cyst nematode parasitism. Molecular Plant-Microbe Interactions **21:** 1622-1634

Hewezi T, Juvale PS, Piya S, Maier TR, Rambani A, Rice JH, Mitchum MG, Davis EL, Hussey RS, Baum TJ (2015) The cyst nematode effector protein 10A07 targets and recruits host posttranslational machinery to mediate its nuclear trafficking and to promote parasitism in Arabidopsis. The Plant Cell **27:** 891-907

Hewezi T, Lane T, Piya S, Rambani A, Rice JH, Staton M (2017) Cyst nematode parasitism induces dynamic changes in the root epigenome. Plant Physiology**:** pp. 01948.02016

Hewezi T, Maier TR, Nettleton D, Baum TJ (2012) The Arabidopsis microRNA396-GRF1/GRF3 regulatory module acts as a developmental regulator in the reprogramming of root cells during cyst nematode infection. Plant physiology **159:** 321-335

Hewezi T, Pantalone V, Bennett M, Stewart CN, Burch-Smith TM (2017) Phytopathogen-induced changes to plant methylomes. Plant Cell Reports**:** 1-7

Hewezi T, Piya S, Richard G, Rice JH (2014) Spatial and temporal expression patterns of auxin response transcription factors in the syncytium induced by the beet cyst nematode *Heterodera schachtii* in Arabidopsis. Molecular plant pathology **15:** 730-736

Hofmann J, El Ashry AEN, Anwar S, Erban A, Kopka J, Grundler F (2010) Metabolic profiling reveals local and systemic responses of host plants to nematode parasitism. The Plant Journal **62:** 1058-1071

Hofmann J, Hess PH, Szakasits D, Blöchl A, Wieczorek K, Daxböck-Horvath S, Bohlmann H, van Bel AJ, Grundler FM (2009) Diversity and activity of sugar transporters in nematode-induced root syncytia. Journal of experimental botany **60:** 3085-3095

Hu Z, Jiang Q, Ni Z, Chen R, Xu S, Zhang H (2013) Analyses of a *Glycine max* degradome library identify microRNA targets and microRNAs that trigger secondary siRNA biogenesis. Journal of integrative plant biology **55:** 160-176

Krueger F, Andrews SR (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. bioinformatics **27:** 1571-1572

Lakhssassi N, Doblas VG, Rosado A, del Valle AE, Posé D, Jimenez AJ, Castillo AG, Valpuesta V, Borsani O, Botella MA (2012) The Arabidopsis tetratricopeptide thioredoxin-like gene family is required for osmotic stress tolerance and male sporogenesis. Plant physiology **158:** 1252-1266

Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature Reviews Genetics **11:** 204-220

Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH, Kim VN (2004) MicroRNA genes are transcribed by RNA polymerase II. The EMBO journal **23:** 4051-4060

Lehmann U (2014) Aberrant DNA methylation of microRNA genes in human breast cancer–a critical appraisal. Cell and tissue research **356:** 657-664

Li X, Wang X, Zhang S, Liu D, Duan Y, Dong W (2012) Identification of soybean microRNAs involved in soybean cyst nematode infection by deep sequencing. PloS one **7:** e39650

Li Z, Lei H, Luo M, Wang Y, Dong L, Ma Y, Liu C, Song W, Wang F, Zhang J (2015) DNA methylation downregulated mir-10b acts as a tumor suppressor in gastric cancer. Gastric cancer **18:** 43-54

Llave C, Xie Z, Kasschau KD, Carrington JC (2002) Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. Science **297:** 2053-2056

Medina C, Rocha M, Magliano M, Ratpopoulo A, Revel B, Marteu N, Magnone V, Lebrigand K, Cabrera J, Barcala M (2017) Characterization of microRNAs from Arabidopsis galls highlights a role for miR159 in the plant response to the root-knot nematode *Meloidogyne incognita*. New Phytologist

Melito S, Heuberger AL, Cook D, Diers BW, MacGuidwin AE, Bent AF (2010) A nematode demographics assay in transgenic roots reveals no significant impacts of the Rhg1 locus LRR-Kinase on soybean cyst nematode resistance. BMC plant biology **10:** 104

Monteys AM, Spengler RM, Wan J, Tecedor L, Lennox KA, Xing Y, Davidson BL (2010) Structure and activity of putative intronic miRNA promoters. Rna **16:** 495-505

Piya S, Kihm C, Rice JH, Baum TJ, Hewezi T (2017) Cooperative regulatory functions of miR858 and MYB83 during cyst nematode parasitism. Plant Physiology **174:** 1897-1912

Pronina IV, Loginov VI, Burdennyy AM, Fridman MV, Senchenko VN, Kazubskaya TP, Kushlinskii NE, Dmitriev AA, Braga EA (2017) DNA methylation contributes to deregulation of 12 cancer-associated microRNAs and breast cancer progression. Gene **604:** 1-8

Rambani A, Rice JH, Liu J, Lane T, Ranjan P, Mazarei M, Pantalone V, Stewart CN, Staton M, Hewezi T (2015) The methylome of soybean roots during the compatible interaction with the soybean cyst nematode. Plant physiology **168:** 1364-1377

Rogers K, Chen X (2013) Biogenesis, turnover, and mode of action of plant microRNAs. The Plant Cell **25:** 2383-2399

Secco D, Wang C, Shou H, Schultz MD, Chiarenza S, Nussaume L, Ecker JR, Whelan J, Lister R (2015) Stress induced gene expression drives transient DNA methylation changes at adjacent repetitive elements. Elife **4:** e09343

Shamimuzzaman M, Vodkin L (2012) Identification of soybean seed developmental stage-specific and tissue-specific miRNA targets by degradome sequencing. BMC genomics **13:** 310

Song Q-X, Liu Y-F, Hu X-Y, Zhang W-K, Ma B, Chen S-Y, Zhang J-S (2011) Identification of miRNAs and their target genes in developing soybean seeds by deep sequencing. BMC plant biology **11:** 5

Song Y, Tian M, Ci D, Zhang D (2015) Methylation of microRNA genes regulates gene expression in bisexual flower development in andromonoecious poplar. Journal of experimental botany **66:** 1891-1905

Sunkar R, Li Y-F, Jagadeeswaran G (2012) Functions of microRNAs in plant stress responses. Trends in plant science **17:** 196-203

Sunkar R, Zhu J-K (2004) Novel and stress-regulated microRNAs and other small RNAs from Arabidopsis. The Plant Cell **16:** 2001-2019

Supek F, Bošnjak M, Škunca N, Šmuc T (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. PloS one **6:** e21800

Suzuki MM, Bird A (2008) DNA methylation landscapes: provocative insights from epigenomics. Nature Reviews Genetics **9:** 465-476

Tauzin AS, Giardina T (2014) Sucrose and invertases, a part of the plant defense response to the biotic stresses. Frontiers in plant science **5**

Wibowo A, Becker C, Marconi G, Durr J, Price J, Hagmann J, Papareddy R, Putra H, Kageyama J, Becker J (2016) Hyperosmotic stress memory in Arabidopsis is mediated by distinct epigenetically labile sites in the genome and is restricted in the male germline by DNA glycosylase activity. Elife **5:** e13546

Xu M, Li Y, Zhang Q, Xu T, Qiu L, Fan Y, Wang L (2014) Novel miRNA and phasiRNA biogenesis networks in soybean roots from two sister lines that are resistant and susceptible to SCN race 4. PloS one **9:** e110051

Xu S, Liu N, Mao W, Hu Q, Wang G, Gong Y (2016) Identification of chilling-responsive microRNAs and their targets in vegetable soybean (*Glycine max* L.). Scientific reports **6**

Yan H, Choi A-j, Lee BH, Ting AH (2011) Identification and functional analysis of epigenetically silenced microRNAs in colorectal cancer cells. PloS one **6:** e20628

Yong-Villalobos L, González-Morales SI, Wrobel K, Gutiérrez-Alanis D, Cervantes-Peréz SA, Hayano-Kanashiro C, Oropeza-Aburto A, Cruz-Ramírez A, Martínez O, Herrera-Estrella L (2015) Methylome analysis reveals an important role for epigenetic changes in the regulation of the Arabidopsis response to phosphate starvation. Proceedings of the National Academy of Sciences **112:** E7293-E7302

Zhao M, Cai C, Zhai J, Lin F, Li L, Shreve J, Thimmapuram J, Hughes TJ, Meyers BC, Ma J (2015) Coordination of microRNAs, phasiRNAs, and NB-LRR Genes in response to a plant pathogen: Insights from analyses of a set of soybean Rps gene near-isogenic lines. The Plant Genome **8**

Zhao W, Li Z, Fan J, Hu C, Yang R, Qi X, Chen H, Zhao F, Wang S (2015) Identification of jasmonic acid-associated microRNAs and characterization of the regulatory roles of the miR319/TCP4 module under root-knot nematode stress in tomato. Journal of experimental botany **66:** 4653-4667

Appendix

Figure 4.1 Absolute DNA methylation level in three contexts across all annotated microRNA in Glycine max cv Williams 82 genome under non infected (CONTROL) and SCN-infected conditions (INFECTED). Global methylation is calculated at base pair level by dividing number of cytosines by total coverage. Promoter region of 2 kb is divided into 60 bins and methylation is averaged across each bin to be plotted as a data point on a line graph. The x-axis is methylation level represented within range from 1-0 and y-axis is the region over which methylation level was calculated..**(A)** Global CG methylation, **(B)** Global CHG methylation, **(C)** Global CHH methylation. **(D)** Percent differential methylation levels in promoters of 28 miRNAs showing differential methylation levels. Values 0 to 100% are hypermethylated and values 0 to -100% are hypomethylated.

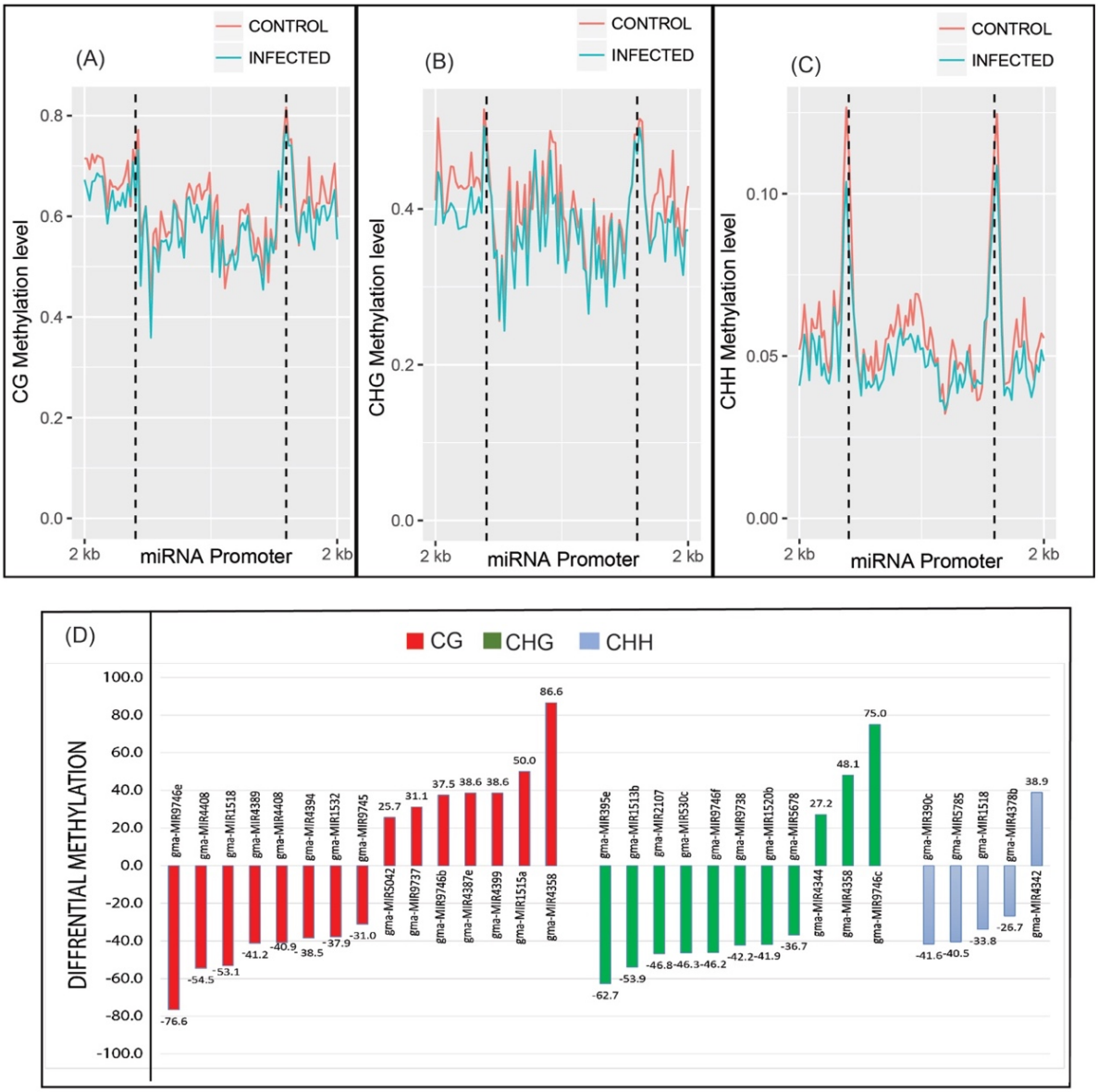

Figure 4.1 continued

Figure 4.2 Functional classification of target genes of differentially methylated miRNAs during infection in Williams82. (A) Significantly enriched GO terms in the category of biological processes for target genes of miRNA with significantly differentially methylated promoter in Glycine max cv williams82 during SCN compatible interaction. GO terms are clustered based on semantic similarity to other GO terms in Uniprot database. The size of the circle is proportional to the abundance of the GO term in our dataset and the color is representative of the log10 of the GO term p-value obtained from soybase GO enrichment tool (higher p-value is red and lower value is blue). **(B)** Summary of annotation of target genes of miRNA with significantly differentially methylated promoter in williams82 during SCN compatible interaction. Annotation was available for 120 genes from soybase and 20 genes did not have any information. Number of genes for each category of annotation is represented by the length of the bar and the color of the bar indicates the miRNA targeting the gene. **(C)** Venn diagram showing the overlap between targets of miRNA with significantly differentially methylated promoter in williams82 during SCN compatible interaction (TARGETS) with genes known to be significantly differentially expressed in the syncytium (DE syncytium). **(D)** Annotation of 23 genes overlapping with differentially expressed genes in syncytium and target gene so differentially methylated miRNA in Williams 82 during infection.

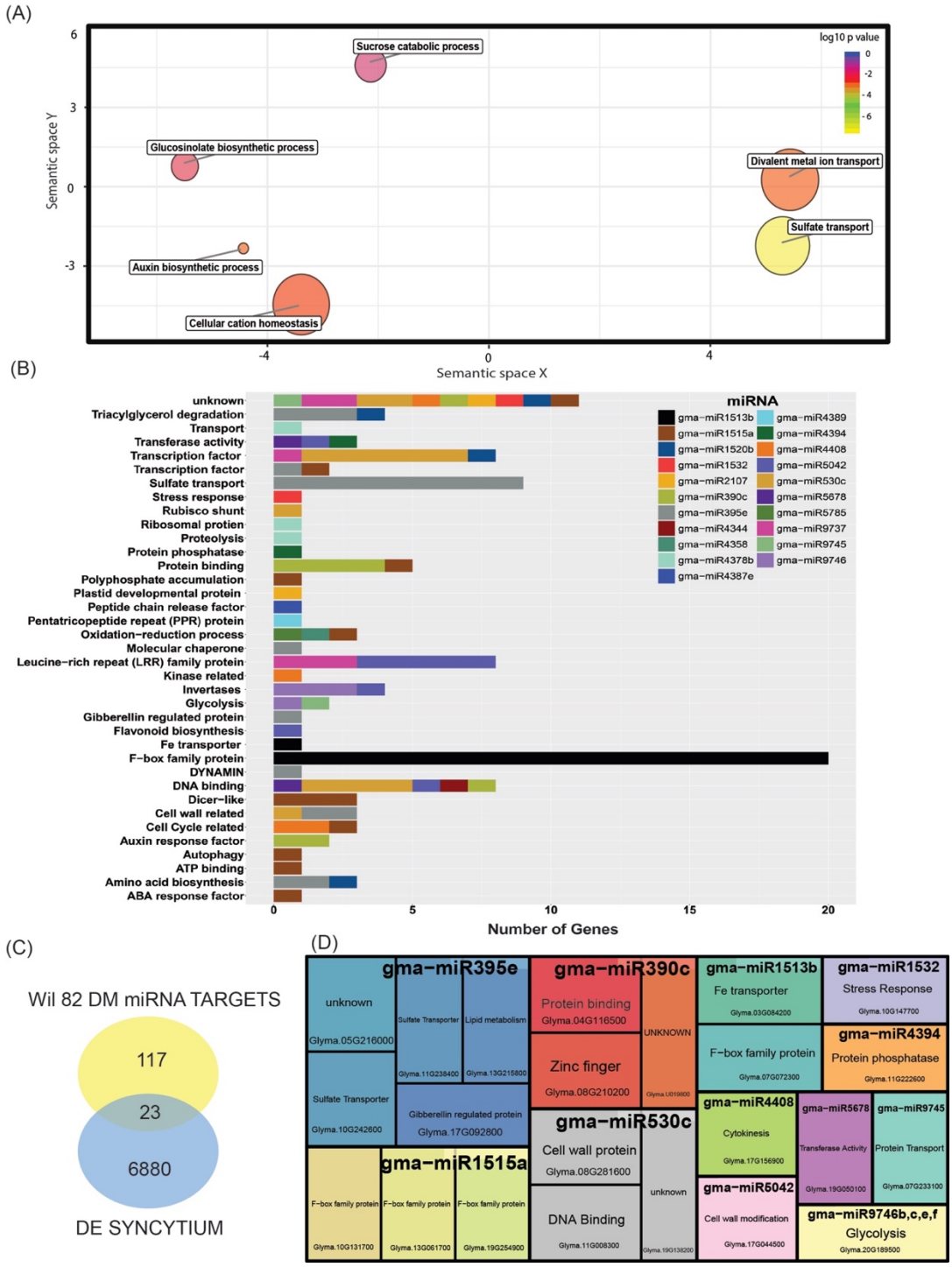

Figure 4.2 continued

Figure 4.3 Comparison of global methylation levels between two near isogenic lines TN09-29 (Resistant) and TN09-16 (Susceptible) in promoter region of all annotated microRNAs under infected and non-infected condition. Global methylation is calculated at base pair level by dividing number of cytosines by total coverage. Promoter region of 2 kb is divided into 60 bins and methylation is averaged across each bin to be plotted as a data point on a line graph. The x-axis is methylation level represented within range from 1-0 and y-axis is region over which methylation level was calculated. **(A)** Global CG methylation **(B)** Global CHG methylation **(C)** Global CHH methylation. Promoter region is 2kb upstream of gene start and absolute methylation is calculated at base pair level by dividing number of cytosine by total coverage.

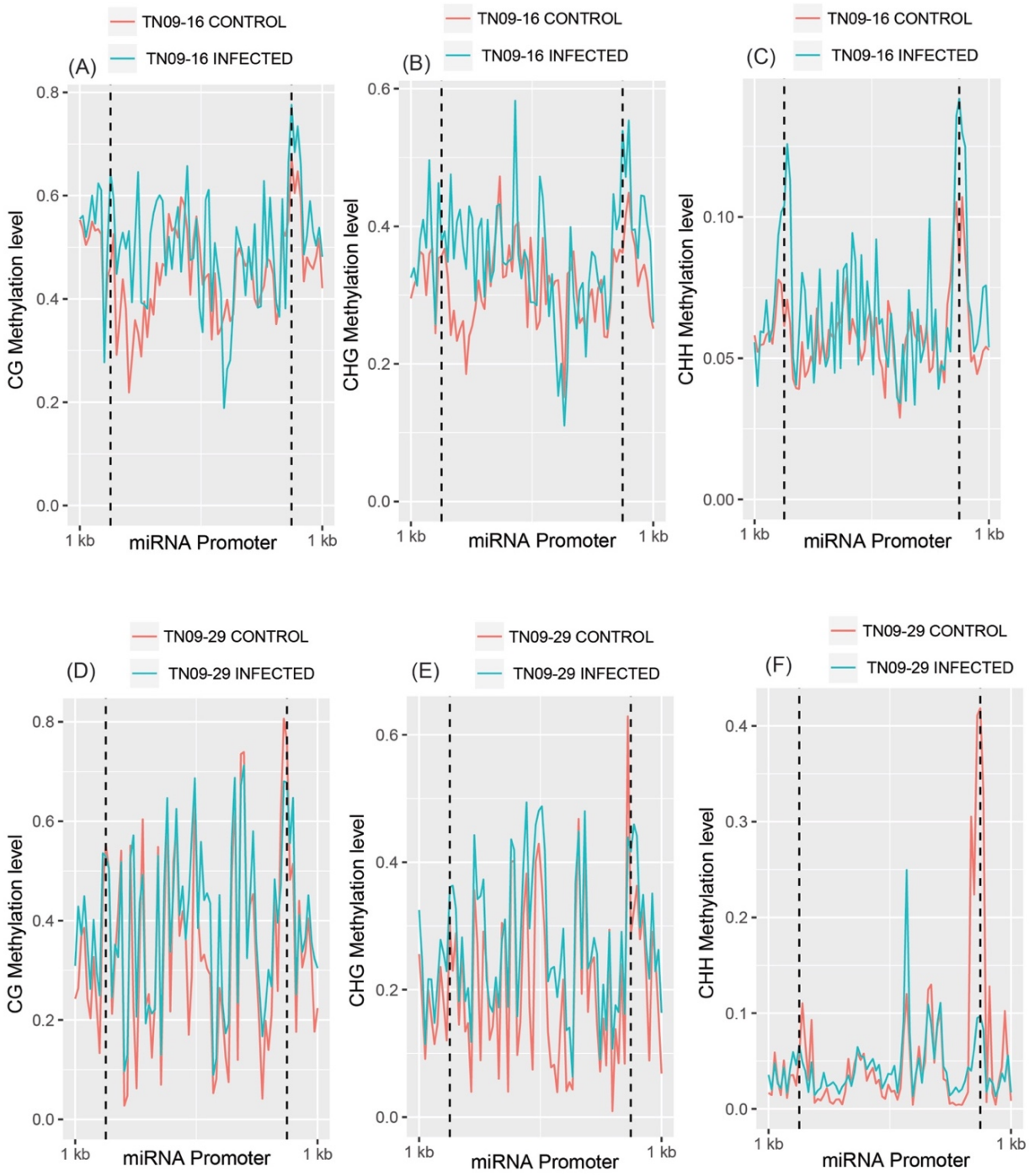

Figure 4.3 continued

Figure 4.4 Differential methylation patterns of the near isogenic lines TN09-29 (Resistant) and TN09-16 (Susceptible) in response to SCN infection. (A) Number of hyper- and hypomethylated microRNAs under SCN infection in CG, CHG, CHH methylation contexts **(B**) Venn diagram showing the Overlap between the differentially methylated miRNAs identified in TN09-29 (TN09-16 (SUS) under SCN infection. **(C)** Comparison of absolute methylation levels of of gma-MIR5032 under SCN infection in the isogenic lines. **(D)** Differential methylation levels for five miRNAs showing differential methylation in both lines.

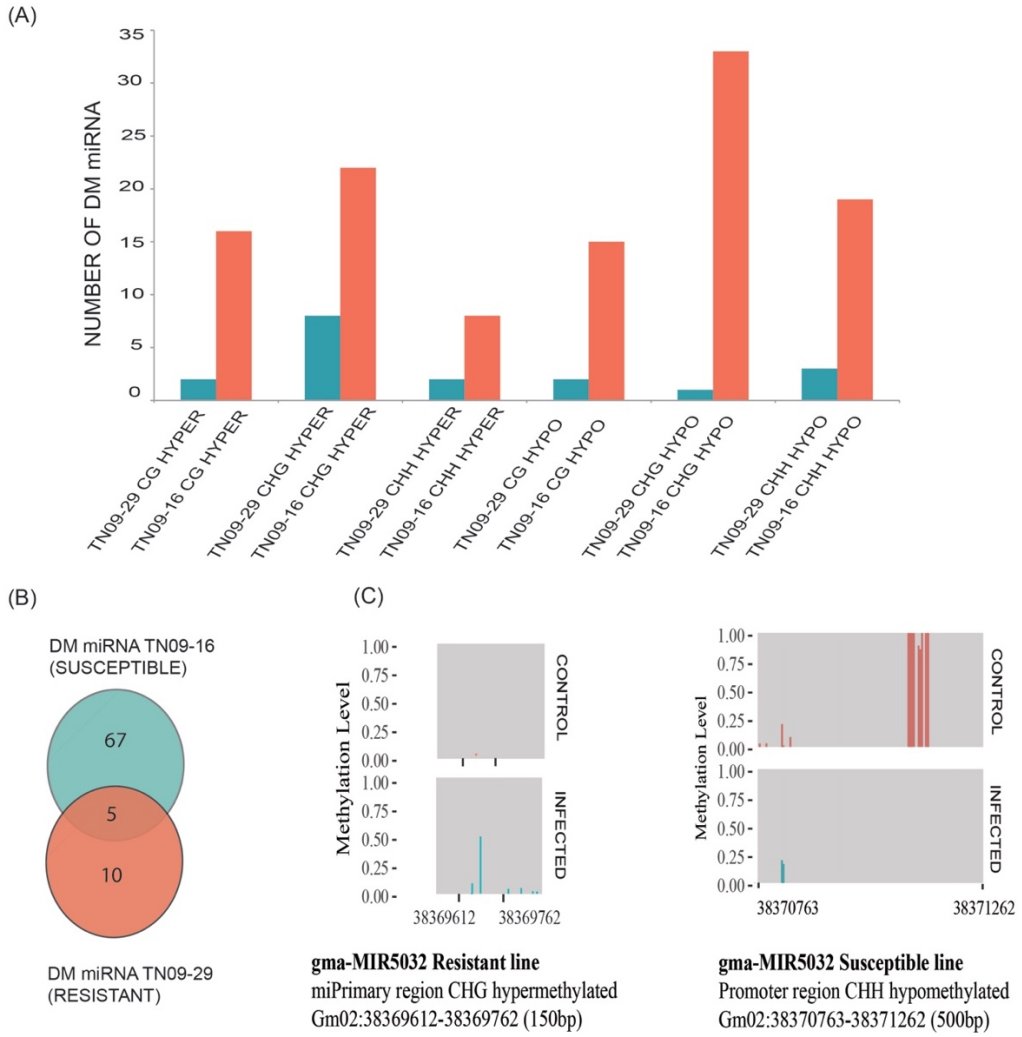

Figure 4.4 continued

Figure 4.5 Functional classification of target genes of miRNAs during infection for compatible and incompatible interaction. (A) Significantly enriched GO terms of target genes of differentially methylated miRNAs identified in the susceptible line (TN09-16) and resistant line (TN09-29) **(B).** GO terms were clustered based on semantic similarity to other GO terms in Uniprot database. The size of the circle is proportional to the abundance of the GO term in our dataset and the color is representative of the log10 of the GO term p-value obtained from soybase GO enrichment tool (higher p-value is red and lower value is blue).

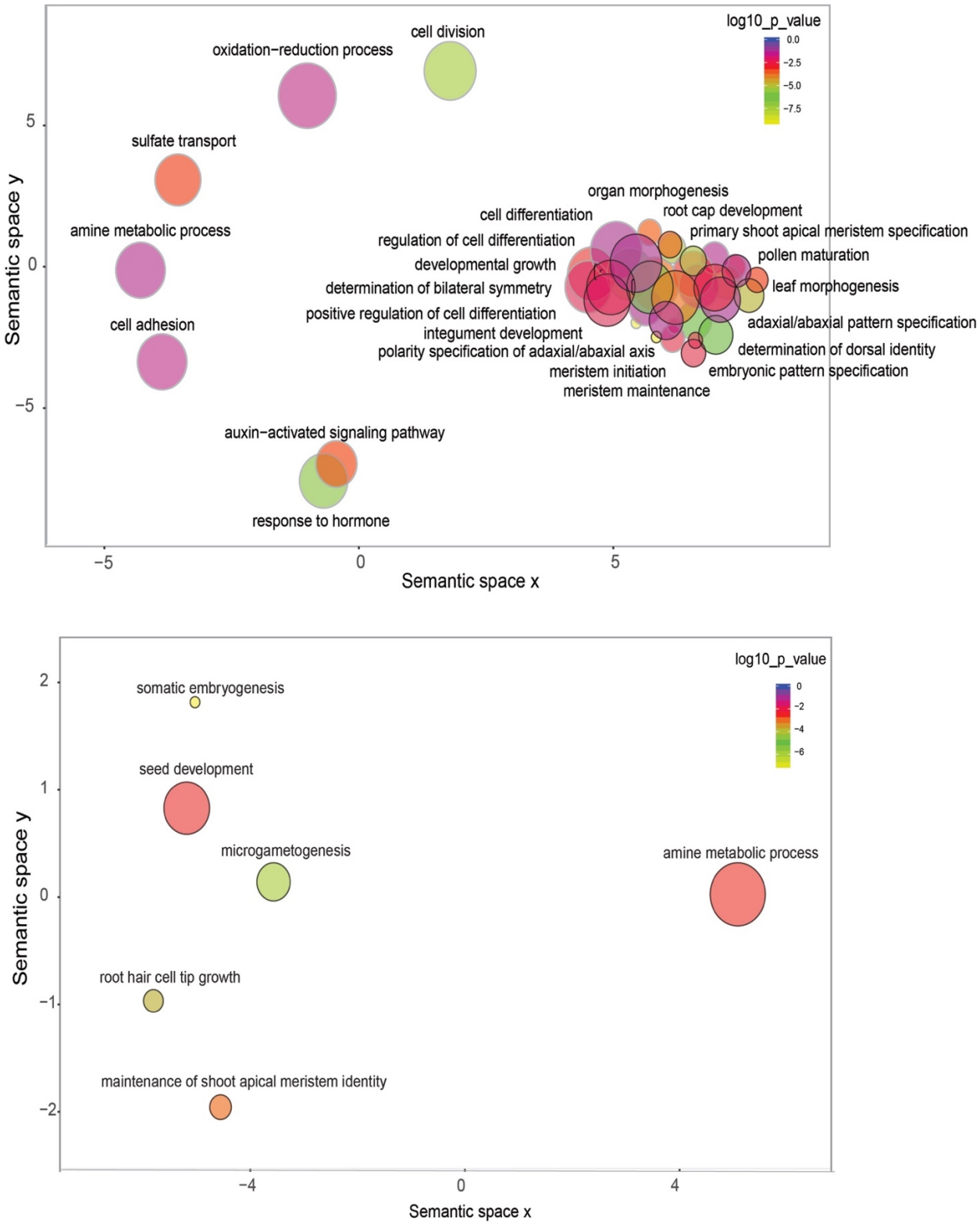

Figure 4.5 continued

Figure 4.6 Overlap between genes differentially expressed in syncytium and target genes of miRNAs differentially methylated during compatible and incompatible interactions with soybean cyst nematode. (A) Overlap between syncytium differentially expressed genes and target genes of significantly differentially methylated miRNAs identified in TN09-16 (A) or TN09-29 (B) in response to SCN infection. **(C)** Annotation of the overlapping genes identified in TN09-16 (C) or TN09-29 (D).

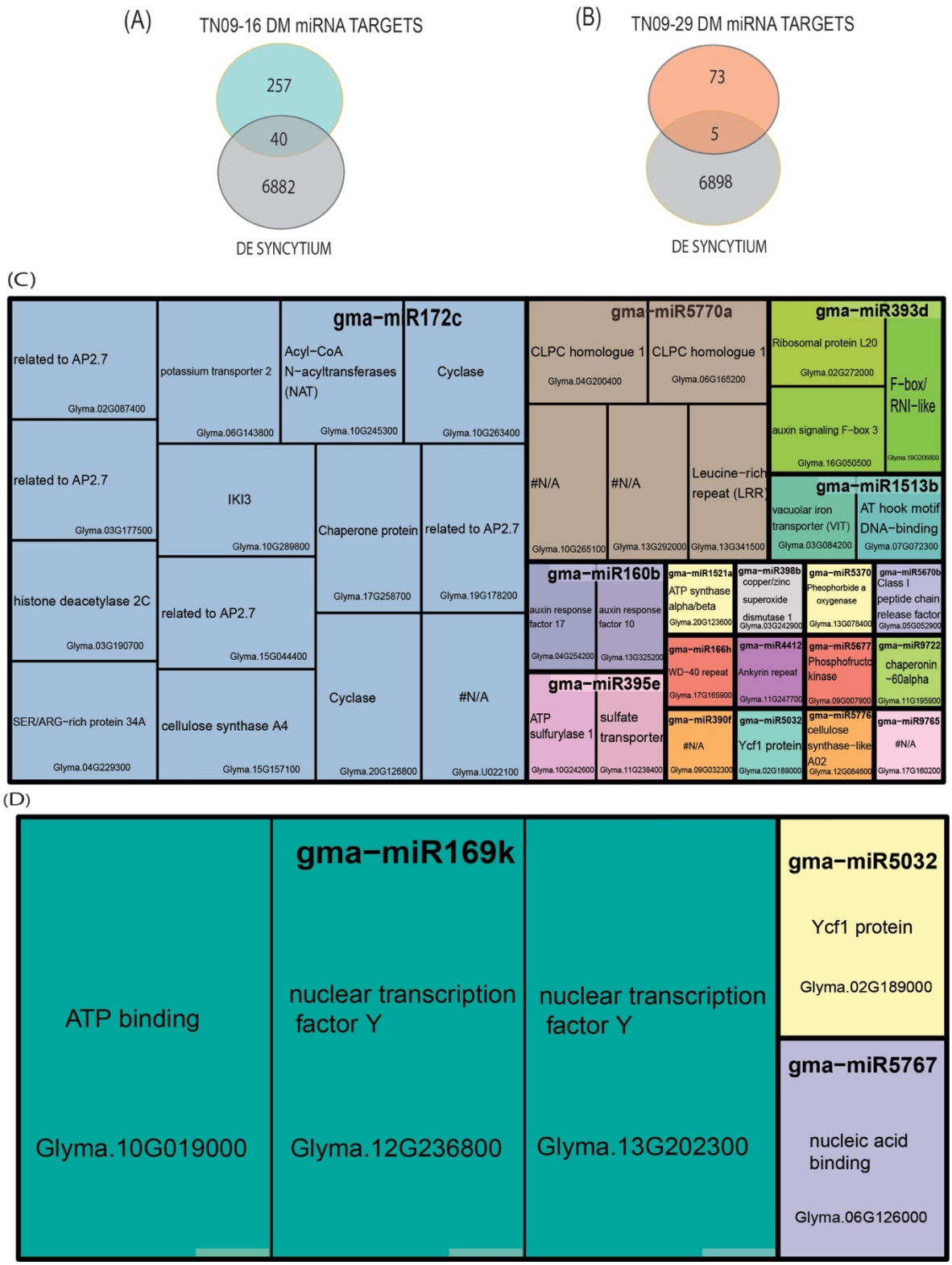

Figure 4.6 continued

Figure 4.7 Functional validation of miR9746 with hairy root assay. (A) Phylogenetic distance tree based on primary sequence similarity between members of miR9746 gene family, six out of nine are zero distance apart. **(B)** Multiple sequence alignment of primary sequences of differentially methylated members of miR9746 gene family, gma-miR9746(b, c ,e, and f). All the members of miR9746gene family have the same mature miRNA sequence. **(C)** Targets of miR9746 predicted based on alignment of longest gene transcript with mature miRNA sequence, using psRNAtarget prediction tool. **(D)** Schematic representation of the construct used for co-expression of green fluorescent protein (GFP) reporter gene and miR9746f. **(E)** Schematic representation of the construct used as control for overexpression of green fluorescent protein (GFP) reporter gene. **(F and G)**. GFP-positive transgenic hairy roots under bright field (F) under GFP filter (G). **(H)** Expression levels of miR9746f in three biological replicates of transgenic hairy roots overexpressing miR9746f, 'Control' represents expression levels in transgenic hairy roots generated using control vector containing GFP marker but not miR9746f gene. qRT-PCR was performed with miRNA9746f specific primers and gene expression values were normalized using soybean ubiquitin 3 gene (GmUBI3). The level of miR9746f expression in the 'Control' transgenic hairy root was arbitrarily set at 1.0. **(I)** Female index of miR9746f overexpressing roots relative to transgenic control roots overexpressing GFP. Data represent average number of cysts obtained from three biological replicates each with 5 plants ± SE.

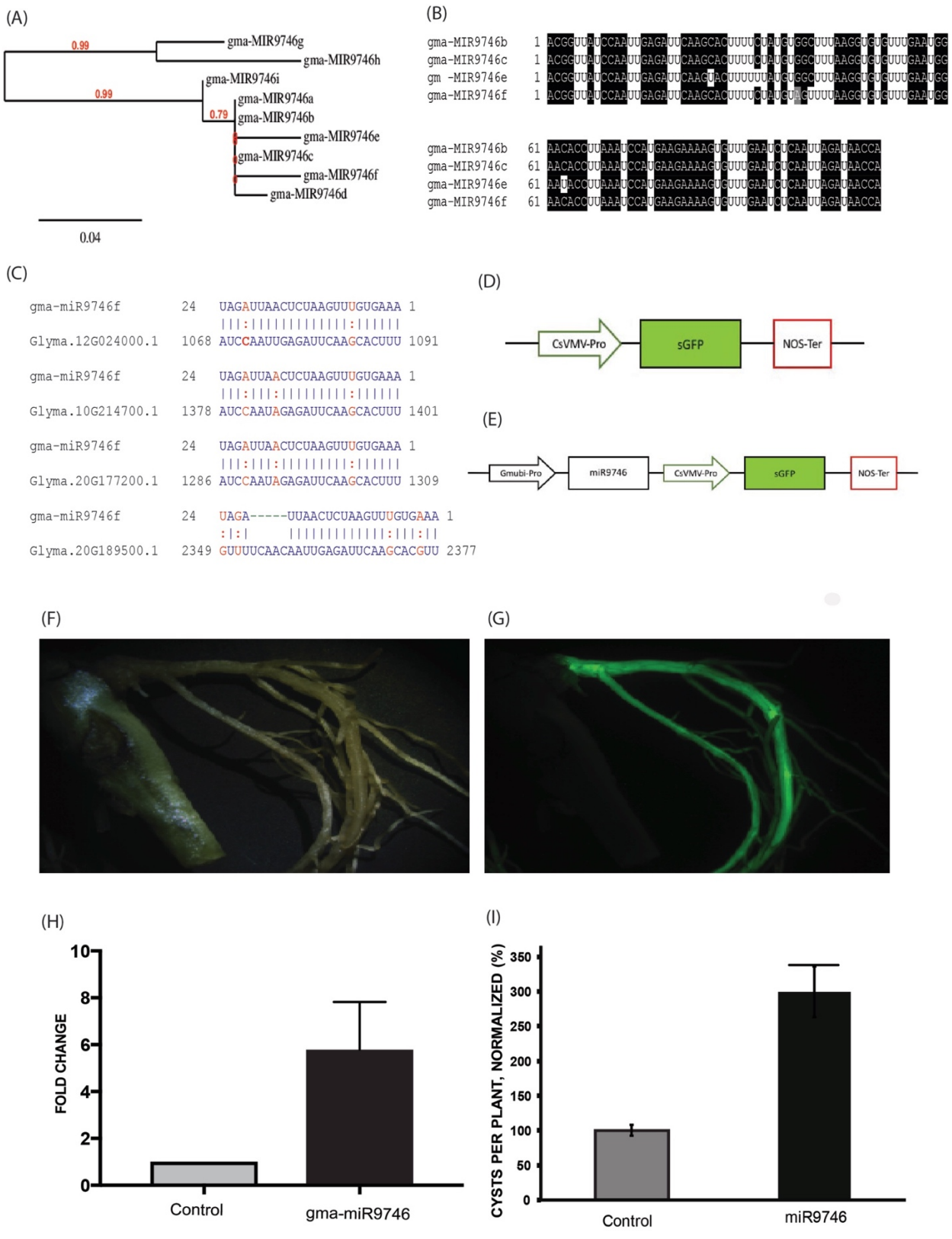

Figure 4.7 continued

Chapter 5: Conclusions

This thesis summarizes projects undertaken to elucidate global DNA methylation changes in soybean as a response to SCN infection. Several experimental and computational approaches were used to achieve this goal. I helped design experiments for generating bisulfite sequence and RNA sequence data from control and infected whole root tissue. I developed a straight forward bioinformatics pipeline for comparison and analysis of whole genome methylation. More specifically, I optimized ideal tile size and sliding window for comparing two methylomes. Methylation of transposable elements in soybean could only be identified after annotation of TEs in the new soybean assembly (Wm82.a2.v1). I developed a methodology for annotating TEs in the new assembly using available soybean TE sequence information from previous assemblies. I built a database with TE annotation, TE coordinates and TE distance from genes for the latest version of soybean assembly (Wm82.a2.v1). In order to identify targets of differentially methylated miRNAs, I computationally predicted targets of all annotated miRNAs in soybean genome. I combined 8 publically available soybean degradome datasets to build a database containing all validated targets of soybean miRNAs.

Our analysis of DNA methylation patterns in the SCN-susceptible cultivar Williams 82 provided the first insight into the role of this important epigenetic mark in regulating gene expression during the compatible interaction between SCN and soybean. Our analysis revealed that loss of DNA methylation occurs to a much higher level than gain of DNA methylation. A recent study of root methylome in Arabidopsis infected with beet cyst nematode, a closely related nematode species to SCN, also showed about 3 folds increase in hypomethylation levels in various sequence contexts relative to hypermethylation (Hewezi et al., 2017). Other plant pathogens including bacteria, fungi and the symbiotic bacteria, Rhizobia, also induce hypomethylation more than hypermethylation during the interaction with their host plants (Dowen et al., 2012; López Sánchez et al., 2016; Satgé et al., 2016).While hypomethylation seems to be a common feature of plant susceptible response to various pathogens, it should be noted that each pathosystem impacts distinct set of genes involved in various cellular processes suited for the nature of the interaction. For example, genes with defense-related functions constitute the main component of the hypomethylated genes during plant-bacterium interactions. In our analysis, we found that hypomethylation impacted

genes involved in various biological processes including cell wall modifications, cellular differentiation, cytoskeleton rearrangement, phytohormone and developmental signaling, epigenetic modifications, gene regulation, and ubiquitination.

One of the main signatures of the soybean methylome induced by SCN infection is the enrichment of CHH and CHG methylation in gene body. This is the contrary to the results reported in other pathosystems where differential DNA methylation in gene body regions occurs mainly in the CG context (Dowen et al., 2012; Yu et al., 2013; López Sánchez et al., 2016). Future studies investigating DNA methylation patterns in response to various biotic and abiotic stress treatments will reveal the extent to which non-CG methylation in the transcribed gene region is unique to SCN infection. Also, investigating the impact of various biotic and abiotic stress factors on soybean methylome will be fundamental to identify common and unique signatures of epigenetic modifications associated with these stress factors. Identifying these epigenetic signatures is the first step towards including epigenetic selection in breeding programs, specifically for agronomic traits with limited genic variations.

The identification of thousands of differentially methylated regions in SCN-infected roots is very interesting and surprising taking into consideration that SCN infection sites (syncytia) represent a very small faction of the root samples used for DNA methylation analysis. This finding underscores the ability of SCN to induce differential DNA methylation in a systemic fashion in root cells that are far from the infection sites. While the molecular mechanism underlying the systemic epigenetic modifications remain elusive, this process may involve epigenetically controlled mobile siRNAs. Consistent with this suggestion, the abundance of siRNAs, specifically the 24-nt siRNA class, was found to associate with the levels of DNA methylation in Arabidopsis roots infected with the sugar beet cyst nematode *Heterodera schachtii* (Hewezi et al., 2017). Epigenetically controlled mobile siRNAs have been previously described in pollen (Slotkin et al., 2009).

One of the major challenges in epigenetic studies is to define whether variations associated with specific phenotypes are due to epigenetic or to genetic variations. Our

Comparison of the methylomes of the resistant and susceptible interactions was facilitated by using highly genetically identical NILs differing in *Rhg4* locus and show opposite responses to SCN infection. This study is the first to compare DNA methylation profiles between the compatible and incompatible interactions in any pathosystem. This analysis provided not only unprecedented insight insights into the biochemical basis of *Rhg4* function in SCN resistance but also revealed the differences in the levels and direction of DNA methylation during the resistant and susceptible interactions of soybean with SCN. The NILs showed not only substantial differences in the level of DNA methylation but also in the direction of DNA methylation. While 50,040 DMRs were identified in the susceptible line in response to SCN infection, only 5,080 DMRs were identified in the resistant lines. Also, loss of DNA methylation was found more dominant than gain of DNA methylation in the susceptible line, but the opposite trend was observed in the resistant line. It remains to be determined whether this contrasted pattern of DNA methylation is conserved in other pathosystems.

We also found that DNA methylation reprograming contributes to gene expression changes in the susceptible line and to a much lesser extend in the resistant line. This could be explained by the differences in the biological processes associated with susceptible and resistant responses. Another striking finding of our analysis is the identification of heritable as well as novel non-parental DMRs in genes with functions related to SCN parasitism of soybean. We identified 56 and 106 novel non-parental DMRs unique to the susceptible and resistant line, respectively. The biological significance of the newly acquired DNA methylation patterns in these lines remains to be determined. Overexpressing these DMR-associated genes in the isogenic lines using transgenic soybean hairy root system will be necessary to reveal their function in SCN resistance/susceptibility. Also, we identified 59 DMRs with differential methylation patterns that were stably inherited from the parents. The methylation differences of these 59 regions between the isogenic lines and the stability of these patterns over 13 generations suggest a role in SCN-soybean interaction, specifically a number of these DMR-associated genes have been previously identified among the syncytium differentially expressed genes.

Because various components of epigenetic mechanisms are mutually interconnected (Zilberman et al., 2008; Cedar and Bergman, 2009; Du et al., 2015), we examined whether DNA methylation occurs in miRNA genes, particularly the promoter region. A significant number of differentially methylated miRNA genes was identified in the SCN-susceptible lines Williams 82 and TN09-016. We found that the target genes of these miRNAs in Williams 82 and TN09-016 share functional similarity despite only 8 miRNA genes were similarly differentially methylated in both lines. Thus, it seems that during the susceptible interaction SCN induces differential DNA methylation in genetically distant lines that impact different miRNA genes with similar regulatory function. Similar to protein-coding genes and transposable elements, our analysis revealed that miRNA genes in the susceptible line are more vulnerable to DNA methylation than the resistant line. While 72 miRNAs were identified as differentially methylated in the susceptible line, only 14 were found in the resistant line. These differentially miRNAs seem to play role in soybean-SCN interaction as many of their targets overlap with our list of syncytium differentially expressed genes.

Functional validation of one (miR9746) of these miRNAs through overexpression in hairy root proved to be necessary for successful SCN parasitism. Remarkably, differential DNA methylation in the primary transcript regions of miRNA genes was detected, specifically in the susceptible lines. Though DNA methylation in promoter region can influence miRNA expression, the significance of DNA methylation in transcribed regions is currently unknown. However, DNA methylation in the primary transcript regions of miRNA genes may prevent transposon insertion and aberrant transcription. In future studies, it would be interesting to determine whether differential DNA methylation in the primary transcript regions is a hallmark of actively transcribed miRNA genes and how this epigenetic mark mediates transcriptional activation or repression of miRNA genes. Collectively, our analysis of DNA methylation at genome level during the susceptible and resistant interactions revealed a key role of this important epigenetic mark in shaping the interaction between soybean and SCN by impacting the expression of high number of protein-coding and miRNA genes as well as transposable elements, specifically those located near genes.

References

Cedar H, Bergman Y (2009) Linking DNA methylation and histone modification: patterns and paradigms. Nature Reviews Genetics **10:** 295-304

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

Du J, Johnson LM, Jacobsen SE, Patel DJ (2015) DNA methylation pathways and their crosstalk with histone methylation. Nature reviews molecular cell biology **16:** 519-532

Hewezi T, Lane T, Piya S, Rambani A, Rice JH, Staton M (2017) Cyst nematode parasitism induces dynamic changes in the root epigenome. Plant Physiology**:** pp. 01948.02016

López Sánchez A, Stassen JH, Furci L, Smith LM, Ton J (2016) The role of DNA (de) methylation in immune responsiveness of Arabidopsis. The Plant Journal **88:** 361-374

Satgé C, Moreau S, Sallet E, Lefort G, Auriac M-C, Remblière C, Cottret L, Gallardo K, Noirot C, Jardinaud M-F (2016) Reprogramming of DNA methylation is critical for nodule development in *Medicago truncatula*. Nature plants **2:** 16166

Slotkin RK, Vaughn M, Borges F, Tanurdžić M, Becker JD, Feijó JA, Martienssen RA (2009) Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell **136:** 461-472

Yu A, Lepère G, Jay F, Wang J, Bapaume L, Wang Y, Abraham A-L, Penterman J, Fischer RL, Voinnet O (2013) Dynamics and biological relevance of DNA demethylation in Arabidopsis antibacterial defense. Proceedings of the National Academy of Sciences **110:** 2389-2394

Zilberman D, Coleman-Derr D, Ballinger T, Henikoff S (2008) Histone H2A. Z and DNA methylation are mutually antagonistic chromatin marks. Nature **456:** 125-129

Vita

Aditi Rambani graduated from the Guru Nanak Dev University, Amritsar, India with B.S (honors) in Botany. She was awarded scholarship and a gold medal for maintaining top academic position at Guru Nanak Dev University for duration of her undergraduate work. She received her Masters in Genetics and Biotechnology at Brigham Young University, Provo, Utah. The MS research project was about inheritance of DNA methylation marks of diploid progenitors by tetraploid Gossypium genome and role of DNA methylation in determining expression bias of duplicated genes. In 2013, she began her PhD at University of Tennessee under guidance of Dr. Tarek Hewezi. The main objectives of PhD dissertation were to elucidate global DNA methylation changes in soybean genome in response to SCN infection.

Funding support for PhD dissertation research came from National Science Foundation, TN-SCORE, Tennessee soybean board and University of Tennessee, Institute of Agriculture (UTIA).