2018

Molecular mechanisms governing plant parasitic nematode signaling and host parasitism

Stacey Nicola Barnes
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/etd

Part of the Agriculture Commons, Genetics Commons, Molecular Biology Commons, and the Plant Pathology Commons

Recommended Citation
https://lib.dr.iastate.edu/etd/16545

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Molecular mechanisms governing plant parasitic nematode signaling and host parasitism

by

Stacey Nicola Barnes

A dissertation submitted to the graduate faculty in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Genetics

Program of Study Committee:
Thomas J. Baum, Major Professor
Diane C. Bassham
Michael J. Kimber
Kan Wang
Steven A. Whitham

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2018

Copyright © Stacey Nicola Barnes, 2018. All rights reserved.
# TABLE OF CONTENTS

**ACKNOWLEDGEMENTS** ..............................................................................................................iv

**ABSTRACT** .......................................................................................................................................v

**CHAPTER 1. GENERAL INTRODUCTION** ................................................................. 1
  Life cycle and infection .................................................................................................................... 1
  Plant immunity ............................................................................................................................... 3
  The mining and functionality of *H. glycines* effector proteins ............................................ 5
  Exosomes .................................................................................................................................. 9
  Small RNAs ............................................................................................................................... 11
  Spliced leaders ............................................................................................................................ 13
  General overview of research chapters ................................................................................... 16
  References ................................................................................................................................... 17

**CHAPTER 2. THE GLAND4 EFFECTOR OF THE PLANT-PARASITIC NEMATODE *HETERODERA SCHACHTII* IS A DNA-BINDING TRANSCRIPTIONAL REPRESSOR** ......................................................... 37
  Abstract ..................................................................................................................................... 37
  Introduction ................................................................................................................................. 38
  Materials and Methods ............................................................................................................. 41
  Results ....................................................................................................................................... 47
  Discussion ................................................................................................................................. 60
  References ................................................................................................................................... 65

**CHAPTER 3. EXOSOMES AND SMALL RNAs: POTENTIAL MESSENGERS INVOLVED IN PLANT NEMATODE PARASITISM** .............. 74
  Abstract ..................................................................................................................................... 74
  Introduction ................................................................................................................................. 75
  Materials and Methods ............................................................................................................. 77
  Results ....................................................................................................................................... 80
  Discussion ................................................................................................................................. 87
  References ................................................................................................................................... 90

**CHAPTER 4. *HETERODERA GLYCINES* UTILIZES HYPERVARIABLE SPLICED LEADERS TO PROMISCUOUSLY SL TRANS-SPLICE A LARGE PORTION OF TRANSCRIPTS** ................................................................. 95
  Abstract ..................................................................................................................................... 95
  Introduction ................................................................................................................................. 96
  Materials and Methods ............................................................................................................. 98
Results ................................................................................................................................. 102
Discussion ............................................................................................................................ 112
References ............................................................................................................................ 115

CHAPTER 5. GENERAL CONCLUSIONS .............................................................................. 120
APPENDIX A. CHAPTER 2 SUPPLEMENTARY INFORMATION ........................................ 125
APPENDIX B. CHAPTER 4 SUPPLEMENTARY INFORMATION ........................................ 132
APPENDIX C. FUNCTIONAL ANALYSIS OF CYST NEMATODE EFFECTOR GLAND 4 TRANSCRIPTIONAL ACTIVITY AND PROTEIN INTERACTORS .................................................................................................................. 146
ACKNOWLEDGEMENTS

I would like to thank my major professor Dr. Thomas Baum and my undergraduate academic advisor Dr. Felicitas Avendaño for their continued mentorship throughout my academic and research career so far. Special thanks also to my committee members and colleagues for your guidance and support.

To the undergraduate students who I had the pleasure of meeting at ISU, thank you for providing me with the opportunity to continually improve my teaching and leadership skills.

To my family, thank you for your unwavering love and support throughout this journey.
ABSTRACT

Pathogen infection of crops causes large-scale annual yield losses for farmers worldwide and hinders global efforts to provide adequate amounts of nutrition for the ever-growing human population. Plant-parasitic nematodes (PPN) are among some of the most devastating pathogens due to their ability to parasitize an expansive variety of agriculturally important crops. In order to identify ways to attenuate PPN infection and limit yield losses it is vital that we increase our understanding of host-PPN interactions. Here we investigate the molecular mechanisms that are occurring both within PPN and at the interface between PPN and their host plants.

Research into PPN-derived secretory proteins, termed effectors is currently the most well studied avenue of research to date. Discoveries made as part of this dissertation make a significant contribution to PPN effector research by identifying the first DNA-binding PPN effector and characterizing its functionality in the model PPN-host system *Heterodera schachtii*, the sugar beet cyst nematode, and *Arabidopsis thaliana*.

The secretion of exosomes and small RNAs from a number of pathogens into their respective hosts, including animal parasitic nematodes, led us to investigate these potential mechanisms in PPN. Exploration into exosome secretion in two different PPN species yielded a small and inconsistent number of exosome-like vesicles. It became apparent that isolation of exosomes from obligate parasites such as PPN poses a number of technical challenges making it impossible to conclude that the release of exosomes is a *bone-fide* parasitic strategy utilized by PPN. High-throughput small RNA
sequencing of *H. glycines* led to the detection of a set of PPN-derived miRNAs that have the potential to target host plant transcripts. Future verification of the functionality of PPN-derived miRNAs within host cells will represent a major breakthrough in our understanding of PPN infection.

Finally, we investigated the role of spliced leader trans-splicing (SLTS) in *Heterodera glycines*, the soybean cyst nematode. The existence of spliced leaders (SL) within the nematode phylum was documented several decades ago but it is still unclear what role SLTS plays within the nematode and how vital this role is to nematode viability. A comprehensive genome and transcriptome-wide SL study was conducted in *H. glycines*, which identified a large set of hypervariable SLs that were collectively found on >2,000 transcripts. The frequent appearance of *H. glycines* SLs on a large number of transcripts makes them an attractive target for future RNAi studies aimed at significantly affecting the viability of *H. glycines*.

Collectively, the research presented in this dissertation furthers our understanding of currently explored parasitic mechanisms and offers plausible insights into promising new avenues of PPN research.
CHAPTER 1. GENERAL INTRODUCTION

_Nematoda_ is the second most diverse metazoan phylum containing more than 23,000 recognized species worldwide, with an estimated one million more that are yet to be discovered. The phylum is comprised of both free-living species and parasitic species that are capable of infecting humans, larger animals and most plants (Blaxter and Koutsovoulos, 2015). Molecular analysis of nematode nuclear small subunit ribosomal RNAs indicates that plant parasitism has evolved independently on at least four occasions (Quist et al., 2015). The _Tylenchida_ order (clade 12) contains multiple economically important plant-parasitic nematodes (PPN) including _Heterodera glycines_, commonly known as the soybean cyst nematode, which causes more than $1 billion in annual yield losses in the U.S. (Allen et al., 2017; Koenning and Wrather, 2010). Increasing our understanding of the parasitic mechanisms and the molecular signaling performed by _H. glycines_ will play a key role in enhancing plant resistance in a variety of economically important crops.

**Life cycle and infection**

_**H. glycines**_ has six life stages; the egg, four juvenile stages (J1-J4), and an adult stage. Hatching of _**H. glycines**_ is increased by the presence of plant exudates within the soil (Masamune et al., 1982). The J1 nematode undergoes its first molt within the egg at which point it emerges from the egg as J2 and begins migration towards the root tissue. When the J2 encounters root tissue it breaks through the cell-wall using a combination of thrusting its anterior mouth spear, known as a stylet, and the secretion
of proteins enriched in cell-wall digesting enzymes. Stylet thrusting and protein secretion continue to assist the nematode as it migrates intracellularly through the cortical cells towards the vascular tissue (Wang et al., 1999; Wyss and Grundler, 1992). The proteins secreted by the infecting nematode are collectively referred to as effectors due to their ability to alter the structure and function of host cells (Hogenhout et al., 2009). The identification and functionality of effectors is currently under investigation and will be described in more detail later in this chapter.

Detection of the vascular tissue terminates nematode migration and initiates the sedentary phase. During the sedentary phase, the nematode methodically probes through the cell wall of a selected cell and extends the stylet to cause invagination of the cell membrane (Wyss and Grundler, 1992). Through the secretion of a complex suite of effectors, many of which have unknown functions to date, the selected cell is transformed into a syncytium (Hussey, 1989). The syncytium is the nematode’s sole source of nutrients for the remainder of its life cycle and is characterized by the dissolution of cell walls to create a large multinucleated structure. Many other morphological changes occur during syncytium formation including loss of the vacuole, alteration in cytoskeletal structure, increased density in the cytoplasm and an increase in ER, ribosomes and golgi (Bleve-Zacheo and Zacheo, 1987; Engler et al., 2004; Jones and Northcot.Dh, 1972; Jones and Northcote, 1972).

Once the syncytium has been established, the nematode use its stylet to alternate between the injection of secretions and ingestion of host cytoplasm (Hussey, 1989). The nematode grows into a pyriform shape while advancing through the juvenile stages. The males emerge outside the root from the J4 cuticle, as a vermiform
nematode in preparation for copulation. The females continue to feed and swell through the J4 and adult stages eventually causing the vulva to protrude through the root tissue ready for copulation (Wyss and Grundler, 1992). Once sexual reproduction has occurred, up to one third of the eggs can be deposited into an egg sac and will hatch soon after, allowing for multiple infection cycles within a single growing season. The remainder of the eggs will stay inside the female body which hardens to form a cyst (Koshy and Swarup, 1971). The eggs can lay dormant inside a hardened cyst for many years until nematode hatching is stimulated by the presence of root exudates in the soil (Fukuzawa et al., 1985; Masamune et al., 1982). The cyst provides protection from adverse conditions such as cold winter temperatures and harmful chemical compounds making it difficult to eradicate *H. glycines* using common farming practices such as crop rotation and pesticide application (Jackson et al., 2005).

**Plant immunity**

As sessile organisms, soybeans, like other plants have evolved an intricate innate immune system to fend off nematodes and other pathogens. Plant immunity can be illustrated as a ‘zigzag’ model (Jones and Dangl, 2006). Briefly, plants possess extracellular receptors collectively known as pattern recognition receptors (PRRs). During infection, PRRs recognize conserved pathogen associated molecular patterns (PAMPs) such as flagellin to establish pattern-triggered immunity (PTI). PPN, including *H. glycines*, produce PAMPs in the form of ascarosides pheromones capable of inducing host defenses and enhancing plant resistance (Manosalva et al., 2015). Pathogens seek to overcome PTI by secreting virulence effectors resulting in host
effector-triggered susceptibility (ETS) (Hogenhout et al., 2009; Jones and Dangl, 2006). In nematodes, the most well-studied strategy to date is the secretion of protein effectors which are deployed by multiple cyst nematode (CN) species, including *H. glycines* (Mitchum et al., 2013). Plants respond by recognizing the presence of certain effectors, either directly or indirectly, through the evolution of resistance (*R*) genes, which typically encode intracellular nucleotide-binding leucine-rich repeat proteins (NB-LRRs). Effector recognition prompts a pro-cell death signal transduction cascade known as effector-triggered immunity (ETI). This is an amplified version of PTI and often manifests itself as the hypersensitive response (HR) resulting in cell death or necrosis (Jones and Dangl, 2006). Secretion of the effector RBP-1 by the potato cyst nematode, *G. pallida*, and recognition by the NB-LRR host protein Gpa2 is thus far the only documented case of effector-*R* gene recognition during CN infection (Sacco et al., 2009). There are other cases in which the presence of a CN effector is shown to trigger the cell death response but the mechanisms behind these responses are unclear (Ali et al., 2015; Liu et al., 2016).

The search for *R* genes in soybean has identified numerous quantitative trait loci (QTL) affiliated with *H. glycines* resistance, with the two major ones *Rhg1* and *Rhg4* (Resistance to *H. glycines*) being identified on chromosome 18 and 8, respectively. The genes that are closely linked to these QTL, three in *Rhg1* and one in *Rhg4*, do not encode for typical NB-LRR resistance proteins suggesting that *H. glycines* resistance is driven by one or more novel mechanisms (Cook et al., 2012; Liu et al., 2017; Liu et al., 2012). The development of resistant cultivars is a key strategy for *H. glycines* management with an estimated yield loss prevention total exceeding $400 million.
between 1975 and 1980 for the resistant cultivar Forrest (Concibido et al., 2004; Kandoth et al., 2011). *H. glycines* juveniles are still able to penetrate the roots of resistant cultivars but they are unable to establish adequate syncytia due to the onset of necrosis (Acedo et al., 1984; Kandoth et al., 2011).

The evolutionary battle between the plant and pathogen continues as a dynamic process in which the pathogen aims to reduce or diversify detected effectors and evolves new evasive effectors to suppress plant defenses. Meanwhile, plant R genes are evolving to enhance their effector recognition capabilities to ensure a robust ETI system (Jones and Dangl, 2006).

**The mining and functionality of *H. glycines* effector proteins**

As mentioned above, PPN secrete a suite of effector proteins in order to successfully parasitize host crops. The majority of PPN effectors accumulate within specialized esophageal gland cells, two subventral and one dorsal, before being secreted through the stylet (Hussey, 1989; Vieira et al., 2011). To date, more than 80 candidate effectors have been identified for *H. glycines* through microaspiration and sequencing of *H. glycines* gland cell RNA (Gao et al., 2001, 2003; Noon et al., 2015; Wang et al., 2001). Individual characterization of a subset of *H. glycines* effectors has been carried out after identification of orthologs in *Heterodera schachtii*, commonly referred to as the sugar beet cyst nematode. Studies conducted using *H. schachtii* can be advantageous for in-depth genomic analysis as *H. schachtii*, unlike *H. glycines*, is uniquely capable of infecting the model host crop *Arabidopsis thaliana* (Gheysen and Fenoll, 2011; Sijmons et al., 1991).
The functional CN effector studies published so far have indicated that the subventral glands play a prominent role during the early stages of infection (Davis et al., 2008). Subventral gland dominance in the early life stages is supported by microscopic observations that show large subventral glands in early juvenile nematodes which gradually reduce in size as the life cycle progresses (Bird, 1967; Endo, 1987). Subventral secretions during migration contain a series of cell-wall modifying effectors including endoglucanases, expansins and pectate lyases. Enzymatic effectors are highly expressed in J2 nematodes and aid in the disruption of root tissue for migration, in some cases by directly binding to the cell wall (Gao et al., 2003; Goellner et al., 2000; Haegeman et al., 2011a; Liu et al., 2016; Popeijus et al., 2000; Qin et al., 2004; Smant et al., 1998; Vanholme et al., 2007; Wang et al., 1999).

During the establishment and expansion of the syncytium, a number of plant proteins relating to cell-wall biogenesis and modification are significantly altered (Goellner et al., 2001; Siddique et al., 2012; Wieczorek et al., 2006). In H. schachtii, the secretion of a cellulose-binding protein effector has been shown to interact directly with the A. thaliana protein pectin methylesterase 3, suggesting that effectors are at least partly responsible for cell-wall modifications in addition to the actual degradation (Hewezi et al., 2008). Interaction between the cellulose-binding protein effector and host pectin methylesterase reduces cell wall pectin methylesterification and accelerates other cell wall enzymatic activities (Hewezi et al., 2008). Another CN effector, 19C07, interacts directly with the auxin influx transporter LAX3 which under normal conditions is responsible for inducing the cell wall-modifying enzyme polygalacturonase as part of lateral root formation (Lee et al., 2011).
The effector 19C07 is part of a larger network of effectors that appear to target the phytohormone auxin, as manipulation of a number of key transcription factors involved in auxin signaling are differentially expressed in the syncytium and surrounding cells (Hewezi et al., 2014; Szakasits et al., 2009). One example is the auxin transcription factor indoleacetic acid-induced 16 (IAA16), which is bound by the nuclear localized CN effector 10A07. IAA16-10A07 binding causes an alteration in expression levels of auxin response factors which are reflective of auxin signaling interference (Hewezi et al., 2015). Manipulations of phytohormone pathways is also evident from the secretion of chorismate mutase effectors by a number of PPN (Gao et al., 2003; Haegeman et al., 2011b; Jones et al., 2003; Lambert et al., 1999; Vanholme et al., 2009). Chorismate mutase is a key enzyme involved in the production of a number of metabolites including auxin and salicylic acid (SA), making chorismate mutase an appealing target for pathogen manipulation (Maeda and Dudareva, 2012). SA in particular is key to plant defense against biotrophic and hemi-biotrophic pathogens so it is not surprising that overexpression of genes involved in the production, regulation and signaling of SA resulted in reduced CN susceptibility (Bari and Jones, 2009; Lin et al., 2016; Matthews et al., 2014). Conversely, a number of SA-deficient mutants exhibited an increased susceptibility to CN infection (Wubben et al., 2008). Increased susceptibility and the downregulation of SA-responsive genes also occurs in plants overexpressing the CN effectors HgGLAND18 and 10A06 (Hewezi et al., 2010; Noon et al., 2016).

CLE signaling in plants represents another target for PPN effectors to alter plant physiology in favor of establishing and expanding syncytia. CLEs are small peptides
involved in a number of aspects of plant growth and development (Yamada and Sawa, 2013). Effectors with sequence similarity to the conserved CLE motif of plants have been identified in a number of PPN (Gao et al., 2003; Gao et al., 2012; Lu et al., 2009; Rutter et al., 2014; Wang et al., 2011; Wang et al., 2001). Evidence exists that PPN CLEs bear functional similarity to plant CLEs and can be post-translationally modified in planta in the same fashion as plant CLE precursors (Chen et al., 2015; Lu et al., 2009; Wang et al., 2010). Susceptibility to nematodes is significantly reduced in plants expressing double stranded RNA that is complementary to PPN CLE genes (Patel et al., 2008).

As observed above, the majority of CN effector studies have focused on the interaction between a CN effector and a host protein. Advances in effector characterization within other plant pathogens such as Xanthomonas spp. suggested that CN effectors are also interacting directly with host DNA (Antony et al., 2010; de Lange et al., 2013; Kay et al., 2007; Kim et al., 2008; Li et al., 2013; Nissan et al., 2012; Nissan et al., 2006; Romer et al., 2007; Streubel et al., 2013; Tian et al., 2014; Wang et al., 2014; Yang et al., 2000). The potential for DNA-binding CN effectors was supported by the identification of novel PPN effectors with predicted nuclear localization signals that share protein identity with histone and helicase DNA-binding regions (Bellaìhiore et al., 2008; Gao et al., 2003; Noon et al., 2015). Chapter 2 of this dissertation is dedicated to the discovery of a novel DNA-binding CN effector, GLAND4, which was functionally characterized as part of this dissertation. Briefly, GLAND4, which has a histone-like motif in the N-terminus, binds specifically to a promoter region within the Arabidopsis genome, termed GLAND4 responsive element 2 (G4RE2). Binding of
GLAND4 to G4RE2 caused a downregulation of the two genes flanking G4RE2, meaning that GLAND4 is the first documented case of a plant pathogenic DNA-binding transcriptional repressor. The downregulated genes that flank G4RE2 are both lipid transfer proteins which have previously been implicated in plant defense (Jung et al., 2003).

**Exosomes**

Breakthroughs in a number of pathosystems have identified other bioactive molecules, aside from effector proteins, that can act as change agents inside the host. One example is the release of exosomes by the animal-parasitic nematode *Heligmosomoides polygyrus* and the human filarial nematode *Brugia malayi*, as well as various other helminths, protozoa and fungi (Buck et al., 2014; Gehrmann et al., 2011; Twu et al., 2013; Wang et al., 2015; Zamanian et al., 2015). Exosomes are secretory vesicles that range from 40-100nm, which are typically used as a form of cell-to-cell communication in multicellular organisms (Raposo and Stoorvogel, 2013; Vlassov et al., 2012). Exosomes are generated through inward budding of the endosomal membrane, to form multi vesicular bodies (MVBs), meaning that their composition is reflective of the proteins, RNA and DNA from their cytoplasm of origin (Trams et al., 1981; Valadi et al., 2007; White et al., 2006).

In animal pathosystems, the internalization of parasite-derived exosomes by host cells is already known to cause alterations in the levels of key cytokines and other factors involved in host immunity, resulting in more successful infection (Silverman et al., 2010; Torrecilhas et al., 2009). An investigation into the role of exosomes in host-
PPN interactions, where little is currently known, is detailed in chapter three of this dissertation. The current lack of knowledge regarding the role of exosomes in PPN and plant-pathogens in general may be partly attributed to a sparse understanding of exosomes within the plant hosts themselves as compared to their animal counterparts (Rutter and Innes, 2017). Transmission electron microscopy studies using plant cells have revealed the presence of MVBs, the vesicles which contain exosomes before their release. The number of plant MVBs, the rate of membrane fusion, and the number of EVs released, can all be altered in response to pathogen infection and treatment with defense hormones (An et al., 2006a; An et al., 2006b; Rutter and Innes, 2017). More significantly, clusters of vesicles have been reported within the extrahaustorial matrix, the region between the fungal and plant cell membranes, during powdery mildew infection. The origin of the EVs is unknown but MVB-like structures from the fungi were observed visually fusing with the fungal plasma membrane making it possible that EVs could be exchanged in either or both directions (Micali et al., 2011).

Alterations in vesicles, collectively termed secretory granules, have been observed in the gland cells of various life stages and species of PPN during plant infection (Baldwin, 1982; Endo, 1987). The secretory granules move anteriorly towards a gland cell valve where they are shown to fuse with the membrane to release their contents (Bird, 1967; Endo, 1984). These studies, which were not specifically designed for granule preservation, noted the presence of granules with varying densities, including ones containing minutely dense spherical bodies (Hussey, 1989). It is possible that a microscopy study aimed at the fixation of PPN granules will better decipher the subpopulations of granules, to accurately assess the potential for
exosome release at the plant-nematode interface. It may also be possible to identify exosomes by incubating parasitic nematodes in media to collect their secretion, as was the case in *B. malayi* (Zamanian et al., 2015).

**Small RNAs**

As noted above, exosomes carry a variety of cargo, including different types of protein, DNA and RNA (Trams et al., 1981; Valadi et al., 2007; White et al., 2006). High-throughput sequencing of pathogenic nematode-derived exosomes from *H. polygyrus* and *B. malayi* revealed the presence of miRNAs (Buck et al., 2014; Zamanian et al., 2015). miRNAs are a class of small RNAs, approximately 21-24 nucleotides in length, which perform sequence-specific post-transcriptional gene silencing (Axtell et al., 2011). Comparisons between the two parasitic nematode datasets yielded a small set of common microRNAs and a large number of species-specific miRNAs (Buck et al., 2014; Zamanian et al., 2015). Amongst the common target genes was lethal-7 (Let-7), which has numerous endogenous targets, the most relevant of those being involved in host immune responses such as macrophage polarization and cytokine expression, making Let-7 an attractive target for manipulation (Banerjee et al., 2013; Schulte et al., 2011). The datasets also included a number of other nematode miRNAs with homology to host miRNAs. Synthetic production and transfection of a subset of these homologous miRNAs from *H. polygyrus* into host cells caused a reduction of reporter gene expression, suggesting that pathogenic nematode miRNAs are able to utilize host miRNA machinery to modulate gene expression. Exosomes from *H. polygyrus* were also found to contain a nematode Ago protein, which implies that exosomes may also
be carrying other components that are necessary for parasite miRNA functionality in the host (Buck et al., 2014).

The presence of small RNAs during PPN infection has not yet been documented but such a mechanism would require cross kingdom functionality of secreted small RNAs. Naturally occurring cross kingdom RNAi is documented in the interaction between the fungal phytopathogen *Botrytis cinerea* and its hosts *Arabidopsis* and tomato. Small RNAs identified in *B. cinerea*, when expressed in the host, were able to utilize host miRNA machinery to suppress genes involved in immunity, resulting in enhanced disease susceptibility (Weiberg et al., 2013). Evidence in support of cross kingdom functionality also exists, when artificially induced, from plant to nematode. Transgenic plants producing self-complementary ‘hairpin’ RNA constructs are able to reduce the expression of PPN target genes, though it is not known whether the PPN gene silencing is due to ingestion of small RNAs or double stranded RNA (Charlton et al., 2010; de Souza et al., 2013; Dinh et al., 2014; Dutta et al., 2015b; Hamamouch et al., 2012; Huang et al., 2006; Kumar et al., 2017; Li et al., 2010a; Li et al., 2010b; Noon et al., 2016; Papolu et al., 2013; Patel et al., 2010; Pogorelko et al., 2016; Sindhu et al., 2009; Xue et al., 2013; Yadav et al., 2006). In an effort to determine the involvement of pathogen-derived small RNAs in host-PPN interactions, a high-throughput small RNA sequencing project is currently underway, the details for which are provided in chapter three of this dissertation.
Spliced leaders

Understanding pathogenic virulence mechanisms such as effector proteins and exosomes deployed by PPN provides a means by which the plant hosts can be modified to defend against PPN and favor plant resistance. Another approach for improving plant resistance is the proactive release of defensive compounds by the plant, which specifically target essential pathways within PPN. As stated above, one example of proactive release is the use of ‘hairpin’ RNA constructs to target essential housekeeping genes (Charlton et al., 2010; de Souza et al., 2013; Dutta et al., 2015a; Kumar et al., 2017; Yadav et al., 2006). A potential drawback relating to the use of RNAi against housekeeping genes is the adverse silencing of genes in non-target organisms. In order to circumvent this problem, unique PPN target molecules and pathways must be discovered.

One such pathway that has the potential to provide unique PPN targets is the pathway required to conduct spliced leader trans-splicing (SLTS) in PPN. Chapter four of this dissertation is therefore dedicated to the discovery and functional analysis of SLTS in *H. glycines*. Splicing is a key step involved in the maturation of mRNA transcripts (Hoskins and Moore, 2012). Cis-splicing is the most widely known splicing mechanism and is defined as the process by which exons that are derived from the same precursor mRNA are ligated together after removal of introns. Less commonly, exons originating from distantly located regions of the genome are ligated together in a process known as trans-splicing. The most common example of trans-splicing involves the addition of a short nucleotide sequence, known as spliced leader (SL), to the front of a multitude of unrelated mRNAs in a process referred to as spliced leader trans-
splicing (SLTS) (Boothroyd and Cross, 1982; De Lange et al., 1984; Van der Ploeg et al., 1982). The SL sequence found at the 5’ end of mRNAs is initially transcribed as part of a larger SL RNA gene. The SL RNA contains a donor splice site that divides the molecule into the 5’ exon-like region termed the SL, and a 3’ intron-like region. The SL intron sequence contains a well-conserved Sm binding motif, which alongside the donor splice site, is essential for the recruitment of splicing machinery (Bruzik et al., 1988; Hannon et al., 1992; Sharp, 1987). In the event of splicing, the SL attaches to the 5’ end of a mRNA while the SL intron sequence forms a Y-structured intermediate with the intron-like sequence from the trans-spliced precursor mRNA and is targeted for degradation (Bektesh and Hirsh, 1988; Murphy et al., 1986).

SLTS was first identified in trypanosomes where all mRNAs are known to receive a SL (Boothroyd and Cross, 1982; Milhausen et al., 1984; Parsons et al., 1984). SLTS has since been found to have evolved independently in select species which have a diverse phylogenetic distribution encompassing nematodes, flatworms, rotifers, chordates, cnidarians, dinoflagellates and arthropods (Douris et al., 2010; Ganot et al., 2004; Krause and Hirsh, 1987; Pouchkina-Stantcheva and TUNNAcliffe, 2005; Rajkovic et al., 1990; Stover and Steele, 2001; Vandenberghhe et al., 2001; Zhang et al., 2007). Evidence for SLTS is present in all nematodes that have been analyzed thus far. The model nematode organism C. elegans is one of the most well studied SLTS organisms to date and possesses two SL sequences, SL1 and SL2, which are present on 70% of transcripts (Allen et al., 2011; Huang and Hirsh, 1989; Krause and Hirsh, 1987; Zorio et al., 1994). A total of 110 SL1 RNA genes are present in the C. elegans genome and mutations that prevent the expression of these genes induce a lethal phenotype.
(Ferguson et al., 1996; Nelson and Honda, 1985). C. elegans SL1 or SL1-like sequences have been discovered in multiple other nematodes, including H. glycines and Globodera spp. but their necessity for viability is yet to be explored in PPN (Bers, 2008; Cotton et al., 2014; Fosu-Nyarko et al., 2016).

In C. elegans >17% of genes are located in operons meaning that they are initially transcribed as polycistronic mRNAs (Allen et al., 2011). It is clear that SLTS plays a key role in resolving C. elegans polycistronic mRNAs so that each coding sequence receives a 5’cap and poly A tail ready for translation (Spieth et al., 1993). Operon resolution, however, is not the only function of SLTS as non-operon genes are also SL trans-spliced in C. elegans. Furthermore, SLs in G. pallida and other SLTS organisms have shown no clear affiliation with operon genes (Cotton et al., 2014; Roy, 2017; Stover and Steele, 2001; Vandenberghe et al., 2001). One proposed function of SLTS is the creation of mRNAs with uniform 5’-ends, often referred to as 5’-end sanitization, as the SL displaces the genic 5’-UTR of pre-mRNA molecules (Davis, 1996). Replacement of the 5’-UTR with a SL is potentially advantageous because the SL limits the accumulation of deleterious sequences that might compromise the mRNA such as out of frame start codons (Hastings, 2005). The 5’-end sanitization hypothesis is supported by a strong trend showing that SLTS occurs very close to, or immediately next to the start codon in many SLTS organisms (Blumenthal and Steward, 1997; Lall et al., 2004). Replacement of the 5’-UTR with a SL may also allow for the incorporation of a common motif that is contained within the SL sequence itself. In Oikopleura dioica, ribosomal genes were found to lack a cis-regulatory 5’ terminal oligopyrimidine (TOP) motif, which is a growth-associated translational regulator commonly found near the
transcriptional start site of ribosomal genes (Amaldi and Pierandreiamaldi, 1990; Avni et al., 1994; Colombo and Fried, 1992; Danks et al., 2015; Tang et al., 2001). Even though \textit{O. dioica} ribosomal genes lack the TOP motif, the motif was identified in 80% of the ribosomal mRNAs, meaning that the mRNAs were trans-spliced to a SL containing a TOP-motif. \textit{C. elegans} ribosomal mRNAs were also commonly SL trans-spliced and although the \textit{C. elegans} SLs do not contain a typical TOP motif, it has been proposed that the TOP motif definition may be too conservative (Danks et al., 2015; Thoreen et al., 2012).

As stated earlier, \textit{C. elegans} utilizes two different SLs. The same two SL sequences were identified in \textit{Globodera rostochiensis} along with 30 other SL sequences, 87% of which were confirmed on transcripts in \textit{G. pallida} (Bers, 2008; Cotton et al., 2014). Interestingly, \textit{H. glycines} also possesses a set of twenty-one SLs that display unique aspects to those identified in \textit{Globodera spp.}, leading to speculation about the role that spliced leader diversification might play in parasitism and nematode viability (Chapter four this dissertation).

**General overview of research chapters**

This dissertation is comprised of three independent research projects (chapters 2-4), all of which aim to investigate the molecular mechanisms that occur at either the interface of host-PPN interactions or within the PPN itself.

As stated above, effector proteins deployed by various CN represent an ongoing area of research for which a greater amount of knowledge is required in order to gain a clearer picture of the host-PPN interface. Chapter two enhances our knowledge of
effectors by providing a detailed functional analysis of the CN effector GLAND4, which to our knowledge is the first documented case of a DNA-binding PPN effector.

Chapter three seeks to assess the potential role that exosomes and small RNAs play in host-PPN interactions. Findings within this chapter highlight some of the technical challenges pertaining to exosome isolation from an obligate parasite. Conclusions from the small RNA sequencing project in *H. glycines* suggest that nematode-derived miRNAs do possess the ability to target host transcripts.

Finally chapter four capitalizes on recently available genome and transcriptome wide data to explore the role of spliced leader trans-splicing with *H. glycines*. This chapter reveals a large set of *H. glycines* SLs that are involved in a wide range of biological processes.

References


Boothroyd, J. C., and G. A. M. Cross, 1982, Transcripts coding for variant surface glycoproteins of Trypanosoma brucei have a short, identical exon at their 5′ end: Gene, v. 20, p. 9.


LAX3 to Control Feeding Site Development: Plant Physiology, v. 155, p. 866-880.


Li, J. R., T. C. Todd, and H. N. Trick, 2010b, Rapid in planta evaluation of root expressed transgenes in chimeric soybean plants: Plant Cell Reports, v. 29, p. 113-123.


Roy, S. W., 2017, Genomic and Transcriptomic Analysis Reveals Spliced Leader Trans-Splicing in Cryptomonads: Genome Biology and Evolution, v. 9, p. 468-473.


Smant, G., J. Stokkermans, Y. T. Yan, J. M. de Boer, T. J. Baum, X. H. Wang, R. S. Hussey, F. J. Gommers, B. Henrissat, E. L. Davis, J. Helder, A. Schots, and J. Bakker, 1998, Endogenous cellulases in animals: Isolation of beta-1,4-


The plant apoplasm is an important recipient compartment for nematode secreted proteins: Journal of Experimental Botany, v. 62, p. 1241-1253.


Yadav, B. C., K. Veluthambi, and K. Subramaniam, 2006, Host-generated double stranded RNA induces RNAi in plant-parasitic nematodes and protects the host from infection: Molecular and Biochemical Parasitology, v. 148, p. 219-222.


CHAPTER 2. THE GLAND4 EFFECTOR OF THE PLANT-PARASITIC NEMATODE *HETERODERA SCHACHTII* IS A DNA-BINDING TRANSCRIPTIONAL REPRESSOR

Modified from a paper published in *Molecular plant pathology*

Stacey N. Barnes, Catherine L. Wram, Melissa G. Mitchum, Thomas J. Baum

Abstract

Cyst nematodes are serious plant pathogens that infect a wide range of economically important crops. One parasitic mechanism employed by cyst nematodes is the production and *in planta* delivery of effector proteins to modify plant cells and suppress defenses to favor parasitism. This study focused on GLAND4, an effector of *Heterodera glycines* and *H. schachtii*, the soybean and sugar beet cyst nematodes, respectively. We showed that GLAND4 is recognized by the plant cellular machinery and is transported to the plant nucleus, an organelle where little is known about plant nematode effector functions. We showed that GLAND4 has DNA-binding ability and repressed reporter gene expression in a plant transcriptional assay. One DNA-fragment that strongly bound to GLAND4 was localized in an *Arabidopsis* chromosomal region associated with the promoters of two lipid transfer protein (*LTP*) genes. These LTPs have known defense functions and are downregulated in the nematode feeding site. When expressed in Arabidopsis, GLAND4 caused downregulation of the two *LTP*
genes in question, which was associated also with increased susceptibility to the plant-pathogenic bacterium *Pseudomonas syringae*. Furthermore, overexpression of one of the *LTP* genes reduced plant susceptibility to *H. schachtii* and *P. syringae*, confirming that *LTP* repression likely suppresses plant defenses. This study made GLAND4 one of a small subset of characterized plant nematode nuclear effectors and identified GLAND4 as the first plant–pathogenic DNA-binding transcriptional repressor with a known *in vivo* target.

**Introduction**

Cyst nematodes are sedentary root endoparasites that infect a wide range of economically important crops including soybean. The soybean cyst nematode, *Heterodera glycines*, is the number one pathogen of soybean, causing over $1 billion in annual yield loss in the U.S (Allen *et al.*, 2017, Koenning & Wrather, 2010). Successful cyst nematode infection involves avoiding or suppressing host defenses while penetrating into the roots of host plants and establishing a feeding site close to the plant vasculature (Endo, 1964). The initial feeding cell is expanded through dissolution of the surrounding cell walls to form a multinucleated structure referred to as a syncytium (Jones & Northcote, 1972, Jones, 1981). Gene expression profiling studies performed on whole roots and microdissected syncytia revealed extensive alterations in host gene expression in response to cyst nematode infection (Hermsmeier *et al.*, 1998, Hofmann *et al.*, 2010, Ithal *et al.*, 2007, Puthoff *et al.*, 2003, Szakasits *et al.*, 2009, Wan *et al.*, 2015). The secretion of effector proteins, defined by
their ability to modify host cell structure and function, are key to the breakdown of plant tissue for migration and the suppression of plant defenses (Hewezi et al., 2010, Hogenhout et al., 2009, Wang et al., 1999). Effectors also play a prominent role in the induction of host cell morphological and physiological changes required for syncytium formation and maintenance (Hewezi & Baum, 2013, Mitchum et al., 2013). Therefore, effectors, directly or indirectly, are responsible for many of the gene expression changes observed in the host plant. However, the molecular mechanisms for triggering gene expression changes remain mostly elusive. The majority of known plant nematode effectors are synthesized in three specialized secretory cells, a single dorsal or two subventral glands, before being secreted through a mouthspine known as the stylet (Hussey, 1989, Vieira et al., 2011). Once inside plant tissue, effectors have been shown to accumulate in the apoplast as well as in various compartments within the plant cell, including the nucleus, where little is currently known about effector functionality (Elling et al., 2007, Hewezi et al., 2010, Hewezi et al., 2015, Huang et al., 2006, Jaouannet et al., 2012, Jones et al., 2009, Lozano-Torres et al., 2012, Vieira et al., 2011).

More than 80 candidate effectors have been identified through isolation and sequencing of RNA within the gland cells of *H. glycines* (Gao et al., 2001, Gao et al., 2003, Noon et al., 2015, Wang et al., 2001). The sequences have been subjected to bioinformatic filtering and verification of transcript accumulation in the gland cells to ensure a high likelihood of secretion during parasitism (Noon et al., 2015). Characterization of a subset of these candidate effectors by identifying effector-
plant protein interactions has revealed functions relating to cell wall modification, antioxidant production, hormone signaling, plant defense suppression and host peptide mimicry (Hamamouch et al., 2012, Hewezi et al., 2008, Hewezi et al., 2010, Hewezi et al., 2015, Lee et al., 2011, Noon et al., 2016, Patel et al., 2010, Pogorelko et al., 2016, Vanholme et al., 2009, Wang et al., 2005). Effectors in other pathogens, most notably *Xanthomonas* spp., are known to interact directly with host DNA to modify host gene transcription (Abhilash et al., 2014, de Lange et al., 2013, Li et al., 2013, Nissan et al., 2006, Yang et al., 2000). Transcription activator-like effectors (TALEs) in *Xanthomas* spp. can bind directly to the promoter of host genes resulting in increased transcription triggered by an activation domain found at the effectors’ C-termini. Altered host expression can result in increased virulence or avirulence depending on the host and bacterial strain (Bing & White, 2004, Kay et al., 2007, Sugio et al., 2007, Szurek et al., 2001). The *Meloidogyne incognita* effector 7H08 is the first documented case of transcriptional activation in plant pathogenic nematodes but the target genes and mechanism of activation are yet to elucidated (Zhang et al., 2015). Despite the massive plant gene expression changes accompanying nematode infections, there are currently no peer-reviewed published records of DNA-binding effectors in plant-parasitic nematodes. This is in part due to a sparse understanding of plant nuclear-targeted nematode effectors. Plant-pathogenic DNA-binding effectors are not limited to gene activation, as the *Xanthomonas* effector XopD binds nonspecifically to DNA and actually represses the transcription of reporter genes (Kim et al., 2008). Identification of a transcriptionally repressing effector and its in vivo targets would represent a
breakthrough discovery in the study of molecular plant–pathogen interactions.

This study details the functional characterization of GLAND4 as the first DNA-binding plant pathogenic transcriptional repressor with a known in vivo target. GLAND4 is a dorsal gland-produced cyst nematode effector, which is recognized and transported to the host nucleus. Sequence analysis suggested GLAND4 is a potential DNA-binding effector due to similarities with histone-like proteins and known transcriptional activators. GLAND4 homologs from H. glycines and the sugar beet cyst nematode H. schachtii share 95% amino acid identity. Functional characterization of GLAND4 was conducted using H. schachtii, which can successfully parasitize the model plant Arabidopsis thaliana (Sijmons et al., 1991). This report shows that GLAND4 is a DNA-binding transcriptional repressor of the genes for two A. thaliana lipid transfer proteins (LTPs), which belong to a multigene family involved in a variety of processes including resistance to biotic and abiotic stressors (Ambrose et al., 2013, Jung et al., 2003, Liu et al., 2015). LTP overexpression is known to increase pathogen resistance (Jung et al., 2005). The downregulation of such genes by GLAND4 is therefore believed to play a key role in suppressing plant defenses.

Materials and Methods

In situ hybridization

Primers designed to the coding sequence of HsGLAND4 were used to generate a 200bp amplicon (Supplemental Table S3). The resulting fragment was used as a template for unidirectional PCR to produce sense or antisense DIG labeled DNA
probes (Roche Life Sciences). The probes were incubated with fixed, permeabilized mixed stage nematodes (de Boer et al., 1998, Gao et al., 2001). Probe detection was performed using alkaline phosphatase conjugated anti-DIG antibody and substrate. A Zeiss Axiovert 100 inverted light microscope was used to visualize the specimens (de Boer et al., 1998, Gao et al., 2001). The results are a reflection of consistent findings in three independent hybridizations.

Subcellular localization

\textit{HsGLAND4}^{Sp} was amplified using the sequence specific primers HsG4\_bait\_F and HgG4\_N1\_R and \textit{HgGLAND4}^{Sp} was amplified using HgG4\_N1\_F and HgG4\_N1\_R (Supplemental Table S3). The PCR products were ligated into the BamHI and EcoRI sites of pSAT6\textit{-}EYFP\textit{-}N1 (Tzfira et al., 2005) at the N terminus of YFP under the control of CaMV35S promoter (35S::\textit{HsGLAND4}^{Sp}\textit{-}YFP). The constructs were confirmed using Sanger sequencing. Transient expression was performed using ballistic bombardment of onion epidermal cells (Bio-Rad) as previously described (Elling et al., 2007). YFP fluorescence was observed using a Zeiss Axiovert 100 inverted light microscope after the cells had been incubated in the dark at 26°C for 24 hours. This process was repeated for three independent experiments.

Genomic SELEX

Genomic SELEX was performed as previously published (Shostak et al., 2004). Briefly, 5ug of Sau3AI (Thermo Fisher Scientific, Waltham, MA, USA) digested A.
*thaliana* genomic DNA was incubated with 2 ug of purified recombinant FLAG-HsGLAND4^{SP}-HIS tagged protein. The protein was immobilized using 30 ul of anti-FLAG affinity matrix (Sigma). After a series of wash steps the DNA was eluted using binding buffer (20 mM HEPES at pH 7.9, 8% glycerol, 10 mM MgCl₂, 10 mM Zn\((C_2H_3O_2)_2\) containing 1 M KCl. The eluted DNA was purified using Qiaquick columns (Qiagen), adaptors were ligated for PCR. The DNA was subjected to 3 subsequent rounds of selection. After the final 3 rounds an aliquot of the eluted DNA was ligated into the NotI site of pBluescriptIIKS+ (Stratagene), transformed into DH5alpha. Sequences were obtained using Sanger sequencing.

**EMSA**

Double stranded DNA fragments were amplified using sequence specific primers (Supplemental Table S3) and Col-0 genomic DNA as a template. The fragments were denatured, labeled with biotin using the Pierce Biotin End DNA Labeling Kit (Thermo Fisher Scientific) and reannealed. Electrophoretic mobility shift assays (EMSA) were conducted using a LightShift Chemiluminescent EMSA Kit (Thermo Fisher Scientific). Nineteen pmol of HsGLAND4^{SP}-HIS or Hs28B03^{SP}-HIS recombinant protein was incubated at room temperature for 20 min with 20 fmol biotin end-labeled double stranded oligonucleotides, 1X binding buffer, 5 mM MgCl₂, 0.05% (w/v) NP-40 and 50 ng/ul (dl-dC) non-specific DNA competitor. In competitor reactions, 4, 6, 8 or 10 pmol of unlabeled oligonucleotides was included. Electrophoresis was performed using 6% Novex TBE 1.0 mm DNA retardation gels for 90 mins at 100 V using
pre-chilled Novex 0.5X TBE running buffer (Thermo Fisher Scientific). The gel was transferred to a positively charged nylon membrane (PerkinElmer, Waltham, MA, USA). Biotin-labeled DNA was detected by immersing the membrane in streptavidin-horseradish peroxidase conjugate and chemiluminescent substrate (Thermo Fisher Scientific), followed by exposure to X-ray film (Research Products International, IL, USA). Quantification of DNA-protein complexes was measured as the ratio of intensity of the shifted band over the total amount loaded as determined using ImageJ Software (NIH).

Quantitative real-time RT-PCR

Total RNA was isolated using miCURY RNA Isolation Kit (Exiqon) according to manufacturer’s instructions. The RNA quantity and quality was assessed on a Thermo Scientific Nanodrop 2000. One microgram of total RNA was treated with DNase I (Thermo Fisher Scientific) and cDNA was synthesized using qScript cDNA SuperMix (Quantabio, Beverly, MA, USA). Two-step qRT-PCR was performed on an iCycler iQ Real-Time PCR machine with reactions containing 12.5ng of template DNA, 300nM of primers and the appropriate amount of iQ SYBR Green Supermix (Bio-Rad). Thermocycler program: 95°C for 3min, 40 cycles of 95°C for 15s and 60°C for 30s followed by establishment of a dissociation curve using the following program: 95°C for 1min, 55°C for 10s and a slow ramp from 55°C to 95°C. A. thaliana and H. schachtii tissue were normalized using Actin as a reference gene; GenBank AY063089.1 and AY443352.1, respectively (Supplemental Table S3). Each qRT-PCR used 3 biological
replicates and 4 technical replicates. The expression levels were calculated using the 2-DDCT method (Livak & Schmittgen, 2001) and statistical differences were determined using a t-test in GraphPad Prism 4.

**Generation of Transgenic Arabidopsis**

A FLAG tag was added to the N-terminal of *HsGLAND4*<sup>Sp</sup>. Fragments for genomic insertion were amplified by PCR using gene specific forward and reverse primers for *HsGLAND4*<sup>Sp</sup>, AT3G22600 and AT3G22620 (Supplemental Table S3). The digested fragments were ligated into the binary vector pBI121 and the sequence was confirmed using Sanger sequencing. The construct was transformed into *Agrobacterium tumefaciens* strain C58 using the freeze-thaw method. The *A. tumefaciens* was transformed into *A. thaliana* using the floral dip method as previously described (Clough & Bent, 1998). The seeds were screened on Murashige and Skoog (MS) medium in the presence of 50mg/L kanamycin to select for transformants. Segregation analysis was conducted on T3 seeds to identify stable transgenic lines before using them in subsequent assays.

**Nematode Infection**

*A. thaliana* were surface sterilized and plated onto modified Knop’s media. After 10 days 250 surface-sterilized J2 *H. schachtii* were inoculated onto the roots of each plant as previously described (Baum *et al.*, 2000). Roots for the time course material were collected into liquid nitrogen after 4, 7 and 14 days post inoculation from both
inoculated and non-inoculated tissue. Seeds for the infection assays were plated into 12-well Falcon tissue culture plates (BD Biosciences) in a randomized fashion. Females were counted after 3 weeks. Statistically significant differences between wild-type (Col-0) and transgenic lines were determined using GraphPad Prism 4 t-test on twenty biological replicates and two independent experiments.

**Bacterial Growth Assays**

*Pst* DC3000 was grown overnight in LB Rif 50 at 30°C 250rpm. The bacteria were adjusted to 1.67 x 10^5 cfu/mL and syringe infiltrated into fully expanded four to five-week-old leaves. Leaf discs were harvested, ground and plated onto LB Rif 50 at 2 hours and 72 hours post inoculation. Raw data for two independent experiments each comprised of four biological and three technical replicates was log transformed and analyzed using GraphPad Prism 4.

**Plant Transcriptional Assays**

To confirm the transcriptional repression of HsGLAND4-sp in plant cells, the yeast Gal4BD-UAS system and the bacterial LexA-LexAop binding sequence were transferred into individual binary T-vectors as previously shown (Tiwari et al., 2004, Zhang et al., 2015). Vector modifications and cloning information are listed in the supplemental methods. Plasmids were then inserted into separate *Agrobacterium tumefaciens* cells, strain GV3101 using the freeze-thaw method. A colony from each construct was grown overnight and the cells were pelleted and resuspended in
sterilized infiltration media at a final optical density of 600nM (OD600) of 0.4. The three sets of cells, the reporter and two regulators, were combined in equal parts into one culture and syringe infiltrated into the entire leaf of 3 individual 6-8-week-old *Nicotiana benthamiana* plants. The leaves were ground on liquid nitrogen after 48 hours. RNA extraction, cDNA and qRT-PCR were performed as stated above. Infiltrated *N. benthamiana* tissue was measured for GFP expression and normalized using Hygromycin as a reference gene (GenBank: FJ905225) as it is driven by CaMV 35S in all modified constructs (Supplemental Table S3). Three independent experiments were conducted and analyzed using GraphPad Prism 4.

**Results**

**Sequence analysis and transcript localization of GLAND4 orthologs**

The candidate effector *GLAND4* was initially discovered after performing microaspiration and sequencing of *H. glycines* esophageal gland cell RNA (Noon et al., 2015). To gain insight into whether GLAND4 is a feasible candidate for functional analysis in the model plant *A. thaliana*, the *GLAND4* sequence was identified in *H. schachtii* cDNA. Unlike *H. glycines*, *H. schachtii* successfully infects *A. thaliana*, allowing for more in-depth functional analyses (Gheysen & Fenoll, 2011, Sijmons et al., 1991). A pairwise sequence alignment of the GLAND4 predicted proteins from *H. glycines* and *H. schachtii* displayed a 95% identity indicating that GLAND4 is a good candidate for functional characterization using *A. thaliana* (Fig. S1, Appendix A). Therefore, all subsequent analyses have utilized *H. schachtii* GLAND4 unless otherwise
stated.

Since most known nematode effectors are synthesized within the specialized secretory gland cells of the nematode, the localization of GLAND4 transcripts within *H. schachtii* was assessed by in-situ hybridization using a labeled probe complementary to GLAND4. A strong signal resulting from successful probe hybridization was detected in the dorsal gland of *H. schachtii* revealing an accumulation of GLAND4 transcripts within this cell type (Fig. 1A). The dorsal gland localization of GLAND4 is consistent with previous findings in *H. glycines* (Noon et al., 2015).

---

**Fig 1. Localization of GLAND4.** (A) In situ hybridization showed a localization of GLAND4 within the dorsal gland of *H. schachtii* upon hybridization to an antisense digoxigenin labeled DNA probe. S = Stylet, DG = dorsal gland cell. Scale bar = 10 μm. **(B)** Subcellular localization revealed that 35S::HsGLAND4<sup>SP</sup>-YFP (top panels) and 35S::HgGLAND4<sup>SP</sup>-YFP (bottom panels) are targeted to the nucleus of plant cells when expressed in onion epidermal cells using particle bombardment. Scale bar = 200 μm. Results shown in A and B are representative of three independent experiments, each with at least 10 biological replicates.
GLAND4 has a 543-nucleotide open reading frame, translating to a predicted protein length of 180 amino acids, which contains an N-terminal signal peptide. The signal peptide is cleaved as part of the protein trafficking process within the nematode and is therefore not included in any subsequent GLAND4 analyses, which is denoted by \(^{-SP}\) in gene construct names. Not counting the signal peptide, GLAND4 exhibits a high percentage of lysine residues at the N-terminus: 36 of the first 86 residues (42%) are lysines, suggesting that GLAND4 may form electrostatic interactions with the negatively charged DNA backbone. The lysine-rich region is followed by a charge-neutral linker region and then an acidic C-terminal region (Fig. S1, Appendix A). The positioning of acidic and hydrophobic amino acids in the C-terminal domain bares resemblance to those found in TALEs and other proteins possessing an acidic activation domain (Table S1, Appendix A).

GLAND4 contains three overlapping predicted bipartite nuclear localization signals ranging from amino acids 37-61 and GLAND4 is predicted to accumulate within the plant nucleus (PSORTII) (Fig. S1, Appendix A). To test this prediction \textit{in planta}, the \textit{GLAND4} \textit{H. schachtii} and \textit{H. glycines} coding sequences were placed under the control of the cauliflower mosaic virus (CaMV) 35S promoter and fused to the yellow fluorescent protein (YFP) coding sequence to create 35S::\textit{HsGLAND4}\(^{-SP}\)-\textit{YFP} and 35S::\textit{HgGLAND4}\(^{-SP}\)-\textit{YFP}. The fusion protein constructs were bombarded into onion epidermal cells where GFP signals were localized exclusively in the plant nucleus (Fig. 1B).

To gain insight into the potential role of GLAND4, a Blastp search was
performed against the NCBI non-redundant database. Aside from hits to *Heterodera* and *Globodera*, which is another CN genus, the N-terminal region of GLAND4, amino acids 36-55, contained a repeat region with similarity to histone proteins from a variety of organisms (Table S2, Appendix A). The distribution of charges, homology to known transcription factors and nuclear localization *in planta* suggested that GLAND4 has the potential to function as a DNA-binding transcription factor-like protein.

**GLAND4 increases pathogen susceptibility in A. thaliana**

To determine the role of GLAND4 during infection, three independent homozygous *A. thaliana* T3 lines (3-10, 5-6 and 6-1) constitutively expressing the *GLAND4* coding sequence were developed to test alterations in pathogen susceptibility. *GLAND4* expression was verified using qRT-PCR (Fig. S2, Appendix A). None of the transgenic lines displayed phenotypic differences or differences in susceptibility to *H. schachtii* as compared to wild-type Columbia-0 (Col-0) (Fig. S3, Appendix A). The fact that *H. schachtii* is already delivering GLAND4 into host plants may account for the observation that the additional GLAND4 production in lines 3-10, 5-6 and 6-1 did not increase susceptibility. However, in order to assess the broader role that GLAND4 might play in defense suppression, the same *GLAND4*-expressing *A. thaliana* lines were infected with the bacterial plant pathogen *Pseudomonas syringae pv tomato* (*Pst* DC3000), which does not use a GLAND4 effector as part of its normal plant infection. Transgenic lines 5-6 and 6-1 displayed an increase in susceptibility
relative to Col-0 demonstrating that GLAND4 is able to influence plant susceptibility to pathogens (Fig. 2).

![Bar graph showing bacterial growth](image)

**Fig 2. GLAND4 increases susceptibility to Pseudomonas syringae.** Bacterial titers of independent homozygous *A. thaliana* T3 lines (3-10, 5-6 and 6-1) constitutively expressing GLAND4<sup>Sc</sup>. *A. thaliana* leaves were measured at 2 hours (0 dpi) and 72 hours (3 dpi) after syringe infiltration with *Pst* DC3000. Differences in bacterial counts were determined in two independent experiments each using four biological and three technical replicates. Bars represent the log transformation of averages. A t-test was used to determine differences between the transgenic susceptibility as compared to Col-0 *A. thaliana*. ±SE, *P* < 0.05.

**GLAND4 is a DNA-binding protein**

Based on the similarity of GLAND4 to histone proteins, genomic Systematic Evolution of Ligands by EXponential enrichment (SELEX) analysis (Shostak *et al.*, 2004) was performed to detect possible DNA-binding properties of GLAND4. The procedure was carried out using recombinant FLAG-GLAND4-HIS protein and enzymatically digested Col-0 genomic DNA. Four rounds of purification and enrichment were
performed (Fig. 3A). Aliquots of the eluted DNA from rounds 2, 3 and 4 were cloned and transformed into *Escherichia coli*, and a total of 170 bacterial colonies were sequenced, yielding 59 unique DNA fragments (Fig. 3B). The fragments ranged between 333 bp and 985 bp in length, with an average length of 655 bp. Of the 20 unique

---

**Fig 3.** Characterization of GLAND4 DNA-binding and confirmation of G4RE2 specificity. (A) Flow chart of the method for in vitro genomic selection. (B) Summary of the sequencing results for the in vitro genomic selection with GLAND4-SP. (C) An EMSA with biotin-labeled DNA probes for three selected fragments (G4RE1, G4RE2 and GRF7). The DNA was incubated without any effector protein, with recombinant GLAND4-SP or with recombinant 28B03-SP, a cytoplasmically localized effector. (D) A competitive binding EMSA using biotin-labeled G4RE2. GLAND4-SP was incubated with 200- and 500-fold molar excess of G4RE2 (lane 3 and 4) or 200-fold GRF7 (lane 5) unlabeled DNA. Results shown are representative of two independent experiments.
fragments identified after round 4, 11 were derived from the *A. thaliana* nuclear genome, with the remaining fragments aligning to chloroplast and mitochondrial DNA.

In an effort to test the hypothesis that GLAND4 functions as a DNA-binding transcription factor, attention was focused on two fragments which mapped to nuclear regions that were less than 3KB from a transcriptional start site (TSS), subsequently referred to as GLAND4 Responsive Elements (G4RE1 and G4RE2). Another noteworthy finding was a fragment from SELEX round 3 that corresponded to the coding region of growth regulating factor 7 (GRF7). This discovery was also scrutinized further because other members of this transcription factor group (GRF1 and GRF3) previously have been shown to be key factors in cell reprogramming during *H. schachtii* infection (Hewezi *et al.*, 2012).

Binding of the G4RE1, G4RE2 and GRF7 fragments to GLAND4 was confirmed using an Electrophoretic Mobility Shift Assay (EMSA) (Fig. 3C). Lanes containing GLAND4 showed a shifted band due to a decreased mobility of labeled DNA, which is indicative of the formation of protein-DNA complexes (Fig. 3C). To ensure that the shifted bands indicating protein-DNA binding are specific to GLAND4, reactions using the same DNA fragments, G4RE1, G4RE2 and GRF7, were performed using the unrelated cytoplasmically-located *H. schachtii* effector 28B03 in place of GLAND4. Lanes containing 28B03 did not display a shift in DNA indicating that binding only occurs in the presence of GLAND4 (Fig. 3C).

The variation in intensity of the shifted bands for lanes containing GLAND4 suggests a stronger affinity for G4RE2 than for G4RE1 and GRF7,
making G4RE2 the focus of further investigation (Fig. 3C). To further quantify such suspected affinity differences, a competitor EMSA was performed with the strongest shifted band (G4RE2) and the weakest shifted band (GRF7). An abundance of unlabeled G4RE2 or GRF7 was used as a competitor probe in reactions that all contained a uniform amount of labeled G4RE2. The HsGLAND4-G4RE2 complex was more strongly affected in the presence of G4RE2 competitor (Fig. 3D lane 3 and 4), than in the presence of GRF7 (Fig. 3D lane 5), indicating that GLAND4 binds more strongly to G4RE2 than to GRF7.

**Repression of genes in the region of G4RE2**

G4RE2 is a 122bp fragment that maps in the UTR of AT3G22600 (NM_113159) approximately 50nt downstream of the transcriptional start site. G4RE2 is also located less than 1KB upstream of the transcriptional start site of AT3G22620 (NM_113160) (Fig. 4A). The two genes associated with G4RE2 are part of a large family of lipid transfer proteins (LTPs), some of which are known to have altered gene expression in response to pathogen infection (Consales et al., 2012, Larroque et al., 2013, Molina & Garciaolmedo, 1993, Qutob et al., 2006). The proximity and orientation of the two LTPs in relation to G4RE2 suggests that the expression of both genes could be affected by the binding of GLAND4 to G4RE2. To test this hypothesis, LTP expression was measured in 3-week old GLAND4 transgenic A. thaliana seedlings from the T3 lines 3-10, 5-6 and 6-1 (Fig. 4B and S4, Appendix A). The expression level of AT3G22630 was also quantified as a control gene due to its downstream location and opposite
orientation, in relation to G4RE2, which suggests that \textit{AT3G22630} is unlikely to be affected by the presence of GLAND4 (Fig. 4A). qRT-PCR showed at least a 2-fold decrease in the mRNA abundance of \textit{AT3G22600} and \textit{AT3G22620} in GLAND4 expression lines when compared to Col-0 (Fig. 4B). The downregulation of both \textit{LTP} genes observed in this study is supported by previous findings that show the downregulation of both genes within the microaspirated contents of the \textit{H. schachtii} feeding site (Szakasits et al., 2009). On the other hand, the expression level of \textit{AT3G22630} was unaffected in the GLAND4 lines (Fig. 4B).

**Fig 4. Genes in the region of G4RE2 are downregulated in GLAND4 transgenic \textit{A. thaliana}.**

(A) Arrangement of the genomic region surrounding G4RE2. Green and red arrowheads indicate the start and stop codons, respectively. (B) Expression of the two genes (\textit{AT3G22600} and \textit{AT3G22620}) closest to G4RE2 and the next nearest gene (\textit{AT3G22630}) were measured in 3-week old transgenic \textit{A. thaliana} expressing \textit{GLAND4}\textsuperscript{3P} using qPCR. Data were normalized to \textit{Actin8} as an internal control and consist of two independent experiments each with three biological replicates and four technical replicates. A t-test was used to identify means that were significantly different from Col-0 \textit{A. thaliana}, set at 1.0, represented by a dashed line. Bars represent the averages ±SE, **\textit{P} < 0.0001.
GLAND4 represses expression of a reporter gene *in planta*

An *in planta* transient expression system involving the *GFP* reporter gene was utilized to investigate the transcriptional repression capabilities of GLAND4. In order to engineer changes in *GFP* expression, the bacterial LexA-responsive element 2xLexAop and the yeast responsive element 6xGAL4UAS were incorporated upstream of the *GFP* start codon (Fig. 5A).

---

**Fig 5.** GLAND4 represses reporter gene expression in planta. **(A)** A schematic representation of the reporter and regulator constructs used in the assay. The bacterial promoter LexAop was fused to the 6xGAL4UAS yeast promoter and inserted upstream of *GFP* in a binary T-vector to create a reporter construct. Regulator constructs were created by fusing the coding sequences of experimental proteins to either LexA or GAL4BD to initiate binding of experimental proteins to the promoter region of the *GFP* reporter gene. All regulator constructs were driven by CaMV35S promoter. **(B)** *GFP* expression was quantified by qRT-PCR in the presence of varying combinations of GAL4BD and LexA regulator fusion constructs. Data were normalized to the expression of *hygromycin* and consist of two independent experiments each with three biological replicates and four technical replicates. Expression levels were calculated relative to the level of *GFP* coinfiltrated with GAL4BD and LexA, which was set at 1.0, represented as a dashed line. A t-test was used to identify means that were significantly different from VP16 activated GFP and LexA without a fusion protein. Bars represent the averages ±SE, *P < 0.01.*
Incorporation of $2xLexAop$ and $6xGAL4UAS$ into the promoter region of $GFP$ allows for the binding and manipulation of $GFP$ expression using the LexA and GAL4 DNA-binding domain (GAL4BD). The ability of proteins to regulate $GFP$ expression can then be tested through the fusion of experimental coding sequences to either LexA or GAL4BD to generate $35S::LexA$-regulator or $35S::GAL4BD$-regulator (Fig. 5A). The basal level of $GFP$ mRNA was established after coinfiltration of the reporter with both binding elements in the absence of experimental coding sequences (Fig. 5B). In order to confirm GLAND4 as a transcriptional repressor, high levels of $GFP$ were then induced using the transactivational domain VP16 from herpes simplex virus fused to the GAL4BD. As anticipated, the coinfiltration of the $35S::GAL4BD$-VP16 and $35S::LexA$ alone resulted in an increase in $GFP$ mRNA (Fig. 5B). To test the role of GLAND4 as a transcriptional repressor, the reporter was coinfiltrated with $35S::GAL4BD$-VP16 and $35S::LexA$-GLAND4. Consistent with the hypothesis that GLAND4 functions as a repressor, the presence of GLAND4 yielded a significant reduction in $GFP$ mRNA (Fig. 5B). To ensure that the $GFP$ mRNA reduction was observed specifically in the presence of GLAND4 and not just the by-product of any protein fused to LexA, a comparably sized portion of the GUS coding sequence was fused to LexA as a negative control. As anticipated, co-infiltration of the $35S::LexA$-GUS fusion with $35S::GAL4BD$-VP16 displayed similar levels of $GFP$ expression to that of the co-infiltration of $35S::LexA$ with $35S::GAL4BD$-VP16 (Fig. 5B). Therefore, the specific reduction of $GFP$ mRNA levels in the presence of GLAND4 is consistent with the hypothesis that GLAND4 functions as a transcriptional repressor.
LTP downregulation occurs during H. schachtii infection

To verify that AT3G22600 and AT3G22620 downregulation is a bone fide strategy during cyst nematode infection, qRT-PCR was performed on H. schachtii-inoculated and non-inoculated Col-0 root tissue. The root tissue was harvested at 4, 7 and 14 days post inoculation (dpi) to represent the gene expression changes incurred through the early infection period. AT3G22600 showed a 2-4-fold downregulation across all time points in H. schachtii-inoculated tissue as compared to mock-inoculated tissue (Fig. 6A and B). AT3G22620 showed a 2-fold and 1.5-fold reduction at 7dpi and 14dpi respectively (Fig. 6A and B). Analysis of GLAND4 mRNA abundance

Fig 6. LTP downregulation in wild-type roots is correlated with GLAND4 expression levels. (A) and (B) qPCR was used to determine the LTP expression levels in the roots of H. schachtii-infected and mock treated wild-type Col-0 A. thaliana at 4, 7 and 14DPI. A t-test was used to compare the means between uninfected (set at 1.0), represented as a dashed line, and infected roots at each time point. The data were normalized to A. thaliana Actin8 as an internal control. (C) qPCR was used to determine GLAND4 expression levels in the roots of H. schachtii-infected A. thaliana at 4, 7 and 14DPI. A t-test was used to compare the means between 4DPI (set at 1.0), represented as a dashed line, and infected roots at subsequent time points. The data were normalized to H. schachtii Actin (AY443352). All results consist of two independent experiments, three biological replicates and four technical replicates. Bars represent the averages ±SE, *P < 0.01 and **P < 0.0001.
in the inoculated root tissue revealed that GLAND4 expression peaks at the time point corresponding to the lowest level of expression for both LTPs (Fig. 6). This correlation of expression further supports the role of GLAND4 in the downregulation of LTPs.

**Fig 7. LTP overexpression reduces pathogen susceptibility.** Independent homozygous A. thaliana T3 lines overexpressing either AT3G22600 (600 2-2, 600 6-2 and 600 9-1) or AT3G22620 (620 6-1 and 620 8-1) were tested for alterations in susceptibility by counting (A) the number of female nematodes present 4 weeks post inoculation with infective second-stage juveniles or (B) bacterial titers of leaves at 2 hours (0 dpi) and 72 hours (3 dpi) after syringe infiltration with Pst DC3000. In (A) Bars represent the averages ±SE and the data consists of two independent experiments, each with 20 plants. In (B) the bars represent the log transformed average ±SE and the data consists of two independent experiments each with four biological and three technical replicates. In both experiments a t-test was used to determine differences in transgenic susceptibility as compared to Col-0 A. thaliana *P < 0.05.

**Overexpression of LTPs decreases pathogen susceptibility in A. thaliana**

In an effort to counteract LTP downregulation and explore the importance of the LTPs during H. schachtii infection, both LTP genes (AT3G22600 and AT3G22620) were individually expressed in A. thaliana under the control of CaMV35S (Fig. S4A and B,
Appendix A), which does not contain a known GLAND4-responsive element. The stable transgenic lines were challenged separately with *H. schachtii* and *Pst* DC3000. Two of the three AT3G22600 overexpressing transgenic lines demonstrated a reduction in *H. schachtii* susceptibility (Fig. 7A) and all transgenic lines for both LTPs demonstrated a decrease in susceptibility to *Pst* DC3000 (Fig. 7B).

**Discussion**

This study analyzed the functional role of the dorsal gland effector GLAND4 of *Heterodera* cyst nematodes. High similarity between the *H. glycines* and *H. schachtii* GLAND4 sequences highlighted GLAND4 as a strong candidate for characterization using the *H. schachtii* – *A. thaliana* model system. The generation of transgenic *A. thaliana* expressing a particular cyst nematode effector has proven useful in understanding the effector’s role during the infection process. In some cases, ectopic expression of an effector resulted in an increased susceptibility to multiple pathogens including *H. schachtii*, bacteria or a RNA virus, as well as alterations in defense marker gene expression (Hewezi et al., 2010, Hewezi et al., 2015, Lee et al., 2011). In this study, however, no change in susceptibility to *H. schachtii* was observed in transgenic plants expressing GLAND4. When considering that we have determined that the function of GLAND4 is to repress genes that play a role in plant defenses, this result is not unforeseeable as *H. schachtii* are likely to already be secreting adequate amounts of GLAND4 required for successful infection. The transgenic plants producing high levels of GLAND4 do show increased susceptibility to *P. syringae*, highlighting the
impact that GLAND4 expression has on plant defenses in the absence of *H. schachtii* derived GLAND4.

GLAND4 is one of many pioneer effectors produced by cyst nematodes that lack similarity to known proteins, and as such, protein structure and function prediction tools cannot provide reliable insight regarding the function of GLAND4 (Gao et al., 2001, Gao et al., 2003, Noon et al., 2015, Wang et al., 2001). A histone-like repeat region within GLAND4 alongside the confirmation of nuclear localization, performed as part of this study, provided reason to test the hypothesis that GLAND4 functions as a DNA-binding protein. Genomic Systematic Evolution of Ligands by EXponential enrichment (SELEX) was used to investigate the potential DNA-binding properties of GLAND4, as this methodology has proven successful for a number of DNA-binding proteins (Chen et al., 2011, Shostak et al., 2004, Whittle et al., 2009). The resulting sequences were initially analyzed for conserved elements that may allow for recognition by GLAND4 but this approach was unsuccessful. It is possible that the long input sequences caused background noise, thereby lowering the chance of finding small conserved motifs. Another possibility is that GLAND4 binds to a number of diverse sequences within the genome. One advantage of using genomic DNA instead of random short nucleotides is that the resulting sequences can be viewed in the context of their position within the organism’s genome. Subsequent analyses focused on three *A. thaliana* genomic regions, two of which were located close to a transcriptional start site (TSS) and one that is within the coding sequence for a member of the GRF family, which has previously been implicated in nematode infection (Hewezi
et al., 2012). The selection of fragments located close to the TSS was, in part, based on the discovery that the plant-pathogenic bacteria *Xanthomonas spp.* secrete TALEs, which are known to bind to the promoter region of genes to activate host transcription (Kay et al., 2007). Subsequent electrophoretic mobility shift assays were able to ascertain that G4RE2 was bound strongly and specifically by GLAND4, making G4RE2 the focus of subsequent analyses.

The role of GLAND4 as the first DNA-binding plant pathogenic transcriptional repressor was evident through the use of an *in planta* transient expression system using constructs similar to those used for characterizing the repression domain of Aux/IAA proteins (Tiwari et al., 2004). GLAND4 was anchored to the promoter region of a reporter gene through fusion to a known DNA-binding protein. There, GLAND4 was able to repress reporter gene expression even when the reporter gene was under the influence of the strong activator VP16. The targeting and downregulation of host genes is a powerful pathogenic strategy as it would imply that GLAND4 is capable of counteracting host genes that may otherwise be highly activated as part of the plant defense response.

The two genes closely associated with G4RE2 in *A. thaliana*, *AT3G22600* and *AT3G22620*, belong to a multigene family of lipid transfer proteins (LTPs) that are present within a large number of plant species (Liu et al., 2015). LTPs are reported to perform a wide variety of functions, the most relevant of which includes defense against biotic and abiotic stressors as well as systemic resistance signaling (Ambrose et al., 2013, Jung et al., 2003, Maldonado *et al.*, 2002, Sohal *et al.*, 1999). Analysis of
LTP promoter regions from oilseed rape, rice and pepper showed an increase in activity after viral, fungal and bacterial attack, respectively (Guiderdoni et al., 2002, Jung et al., 2005, Sohal et al., 1999). The induction of multiple LTPs, including AT3G22600 and AT3G22620 has been observed in response to wounding or pathogen attack (Consales et al., 2012, Larroque et al., 2013, Molina & Garciaolmedo, 1993, Qutob et al., 2006). AT3G22600 appears to be involved in pathogen-associated molecular pattern-triggered immunity because this gene is highly upregulated in response to Phytophthora parasitica cellulose-binding elicitor protein and oomycete-derived Nep1-like proteins (Larroque et al., 2013, Qutob et al., 2006). In the current study, AT3G22600 and AT3G22620 were both downregulated in GLAND4-expressing A. thaliana, when compared to wild-type Col-0, confirming that GLAND4 acts as transcriptional repressor, likely by binding to G4RE2 in vivo. The same two LTPs, which are already known to be downregulated within the syncytia, were also found to be downregulated in Col-0 root tissue during the course of cyst nematode infection, verifying that downregulation of AT3G22600 and AT3G22620 is a bona fide occurrence during H. schachtii infection (Szakasits et al., 2009). Furthermore, the findings from the reporter gene assay alongside the downregulation of both AT3G22600 and AT3G22620 during H. schachtii infection indicate that HsGLAND4 can repress transcription by binding either up or downstream of the transcriptional start site. A correlation in GLAND4 expression and the expression of AT3G22600 and At3G22620 was also observed in the Col-0 time-course material, as GLAND4 expression within the infecting nematodes peaked when LTP expression was at its lowest. Downregulation
of AT3G22620 has previously been documented in response to oral secretions from the herbivorous insects Pieris brassicae and Spodoptera littoralis (Consales et al., 2012).

To ascertain how important LTP downregulation is during H. schachtii infection, the LTP genes in question were individually overexpressed in A. thaliana. Elevated AT3G22600 expression reduced H. schachtii susceptibility in two lines that expressed the gene around 35-fold higher than Col-0. Overexpression of AT3G22620 did not alter H. schachtii susceptibility, indicating that downregulation of AT3G22620 could be a by-product of targeting AT3G22600. The hypothesis that AT3G22600 is the primary target is also supported by previous findings that showed AT3G22600 to be among the top 25 most downregulated genes within the H. schachtii feeding site (Szakasits et al., 2009). It is also possible that the AT3G22620 defense mechanism deployed against H. schachtii is limited by its dependence on a cofactor that was not increased in these transgenic lines. All transgenic lines for both LTPs displayed an enhanced resistance to Pst DC3000, which supports previous findings that overexpression of a pepper LTP (CALTP1) in A. thaliana enhanced resistance to Pst DC3000 (Jung et al., 2005).

GLAND4 is the first plant-parasitic nematode effector to join a very small set of DNA-binding plant pathogen effectors, which includes effectors from Agrobacterium tumefaciens (Wang et al., 2014), Xanthomonas spp. (Kay et al., 2007, Sugio et al., 2007),Ralstonia solanacearum (de Lange et al., 2013, Li et al., 2013) and Pantoea agglomerans (Nissan et al., 2006, Nissan et al., 2012). TALEs secreted by Xanthomonas spp. are perhaps the most widely known DNA-binding plant pathogen effectors due to
their ability to function as a tool for genome editing in the broader scientific community (Cox et al., 2015, Scott et al., 2014). As mentioned earlier, TALEs are able to bind to the promoter region of host genes and activate transcription. Here we describe that GLAND4 is capable of binding to DNA and repressing transcription of two genes that function in plant defense. This is a unique finding within plant-microbe interactions and is, to our knowledge, the only reported example so far of a DNA-binding plant pathogen effector with the ability to repress transcription of a known in vivo target.

References


the Xanthomonas type III effector AvrBs3: protein domains involved in
transcriptional activation and the interaction with nuclear import receptors from
pepper. Plant Journal, 26, 523-534.


vectors: a modular series of plasmids for autofluorescent protein tagging and
expression of multiple genes in plants. Plant Molecular Biology, 57, 503-516.

Structural and functional investigation of a secreted chorismate mutase from the
plant-parasitic nematode Heterodera schachtii in the context of related enzymes
from diverse origins. Molecular Plant Pathology, 10, 189-200.

plant apoplast is an important recipient compartment for nematode secreted

gene expression profiling revealed genes and pathways potentially involved in
regulating interactions of soybean with cyst nematode (Heterodera glycines
Ichinohe). Bmc Genomics, 16.

peptide-selection of cDNA cloned directly from the esophageal gland cells of the
soybean cyst nematode Heterodera glycines. Molecular Plant-Microbe Interactions,
14, 536-544.

localization of a beta-1,4-endoglucanase secreted by Heterodera glycines. Molecular
Plant-Microbe Interactions, 12, 64-67.

gene from a plant-parasitic nematode with function similar to CLAVATA3/ESR
(CLE) of Arabidopsis thaliana. Molecular Plant Pathology, 6, 187-191.


CHAPTER 3. EXOSOMES AND SMALL RNAs: POTENTIAL MESSENGERS INVOLVED IN PLANT NEMATODE PARASITISM

Modified from a manuscript to be submitted for publication to the *Journal of Nematology*

Stacey N. Barnes, Thomas R. Maier, Rick E. Masonbrink, Tracey P. Stewart,

Catherine L. Wram, Andrew J. Severin, Thomas J. Baum

Abstract

Plant-parasitic nematodes (PPN) are serious plant pathogens that significantly reduce the yield of numerous economically important crops. To date, a large majority of molecular PPN research has focused on the identification and understanding of proteins, termed effectors, which are secreted by PPN into host plant cells. Recent advances in other areas of parasitology have highlighted key roles for parasite-derived vesicles referred to as exosomes and the small RNAs contained within exosomes as host modifiers during parasitism. This study focused on the identification of exosomes in the PPN species *H. glycines* and *M. incognita*, as well as extensive *H. glycines* small RNA sequencing. Experimentation using current exosome isolation methods recovered the presence of a low quantity of inconsistent exosome-like structures within PPN secretions. Low quantity isolation hindered efforts to identify small RNAs within PPN secretions leading to the strategy of whole nematode *H. glycines* small RNA sequencing. The identification of *H. glycines* miRNAs and their subsequent predictions
within the host transcriptome G. max has revealed the potential for parasite-derived miRNAs as targets for host plant transcripts.

**Introduction**

Cell-to-cell communication through direct connections or via the extracellular space is an essential process performed by multicellular organisms. One cell communication strategy is the release of vesicles including exosomes, ectosomes, shedding vesicles and apoptotic blebs (Raposo and Stoorvogel, 2013; Vlassov et al., 2012). Recently, exosomes have been the most intensely studied of all microvesicles (Keller et al., 2006; Vlassov et al., 2012). Current evidence suggests that exosomes are formed through inward budding from the membrane of multi vesicular bodies (MVB) to generate vesicles inside the lumen of MVBs (Hurley, 2008; White et al., 2006). Upon fusion of the MVBs with the plasma membrane, the internal 40-100nm vesicles are released as exosomes into the extracellular space (Pant et al., 2012; Raposo et al., 1996; Raposo and Stoorvogel, 2013). Due to the formation of exosomes through inward budding of the MVB, exosomal contents contain components of their cytoplasm of origin including proteins, RNA and DNA (Trams et al., 1981; Valadi et al., 2007).

In recent years, a number of parasites including various protozoa, fungi and helminths have been shown to release exosomes during parasitism (Buck et al., 2014; Gehrmann et al., 2011; Twu et al., 2013; Wang et al., 2015; Zamanian et al., 2015). Parasite-driven exosome release plays a prominent role in host immune manipulation
as exosomes released by the helminths *Leishmania donovani* and *Trypanosoma cruzi* cause alterations in host defensive proteins to enhance infection (Coakley et al., 2015; Murray et al., 2002; Silverman et al., 2010; Torrecilhas et al., 2009). Also, the animal-parasitic nematode *Heligmosomoides polygyrus* represses host defensive molecules through the release of exosomes causing the suppression of immunity in host cells (Buck et al., 2014).

Sequencing the nucleic acid contents of parasitic nematode exosomes from *H. polygyrus* and *Brugia malayi* revealed the presence of mature miRNAs with homology to host miRNAs (Buck et al., 2014; Zamanian et al., 2015). Similarly, nematode miRNAs from *Dirofilaria immitis* and *Onchocerca volvulus* were identified in the serum of their hosts demonstrating that exosomes and small RNA release are widespread mechanisms within *Nematoda* (Tritten et al., 2014). It is currently unknown whether the release of exosomes or small RNAs is utilized during plant-parasitic nematode (PPN) infection. In the case of other plant pathogens, naturally occurring cross-kingdom RNAi that uses host machinery is known to occur between the fungal plant pathogen *Botrytis cinerea* in both tomato and *Arabidopsis* (Weiberg et al., 2013).

PPN infection causes billions of dollars in yield losses every year through the redirection of nutrients away from the growth and development of their hosts. The sedentary nature of cyst nematodes (CN) (*Heterodera* spp. and *Globodera* spp.), and root-knot nematodes (RKN) (*Meloidogyne* spp.), requires that they form an intricate relationship with their plant hosts, while finding ways to suppress or circumvent the plant defense system. To date, the characterization of host-PPN interactions has focused largely on the study of stylet-secreted effector proteins and their ability to
trigger reprogramming of host cells (Davis et al., 2008; Hewezi and Baum, 2013; Mitchum et al., 2013; Quentin et al., 2013). A large majority of effectors in both CN and RKN are localized within specialized esophageal gland cells; two subventral and one dorsal. Exploration into exosome trafficking and small RNA delivery from PPN gland cells will create a clearer picture of how feeding sites are formed and maintained during infection.

Previous studies have identified effector proteins by using chemical stimulants to induce esophageal gland cell secretions from PPN (Bellafiore et al., 2008; Davis et al., 1994; Goverse et al., 1994; Jaubert et al., 2002; McClure and Vonmende, 1987). In this study, similar induction methods were used to examine the secretions of *H. glycines* and *M. incognita* for the presence of exosomes. Whole nematode small RNA libraries of *H. glycines* were also generated in an effort to identify PPN-derived miRNAs with the potential to target host transcripts.

**Materials and Methods**

**Secretion Induction**

*H. glycines* and *M. incognita* were retrieved from hatch chambers and washed 3 times in MES-buffered water. *M. incognita* were treated with 0.4% resorcinol (McClure and Vonmende, 1987). *H. glycines* were treated with a varying concentration of Quipazine maleate salt: 0.1mg/ml, 0.25mg/ml, 0.5mg/ml or 1mg/ml. Both species were incubated for 4 hours at room temperature. Visualization of secretions was carried out by adding 0.01% aqueous Coomassie Brilliant Blue G-250 to a small aliquot of
nematodes at the same time as the inducing chemical. Secretions were visualized using a Zeiss Axiovert 100 inverted light microscope.

**Exosome Isolation**

Freshly collected secretions were filtered using a 0.22μm syringe. Initially, exosomes were isolated using ExoQuick (System Biosciences) according to manufacturer’s instructions and resuspended in 30μl of H₂O. Later, the exosomes were isolated as described in (Rani et al., 2011). Briefly, the filtrate was centrifuged at 2,000xg at 4°C for 30mins to remove debris. The supernatant was transferred and the exosomes were pelleted by ultracentrifugation at 110,000xg at 4°C for 2 hours using a SW28 rotor (Beckman Coulter). The exosomes were washed with PBS and ultracentrifuged at 110,000xg at 4°C for 70 min and resuspended in 50μl of PBS.

**Electron Microscopy**

Freshly isolated exosome samples were placed onto mesh carbon film copper grids and contrasted with 2% uranyl acetate. Once the grid was dried a JEOL 200kV JEN 2100 transmission electron microscope (Japan Electron Optics Laboratory, Peabody, MA) was used to image the samples.

**RNA Isolation and Analysis**

Total RNA was isolated from purified exosomes or liquid nitrogen-ground nematode tissue using the Biofluids miCURY RNA Isolation Kit (Exiqon) according to manufacturer’s instructions. The RNA quality and size distribution was assessed using
a Small RNA chip on an Agilent 2100 Bioanalyzer (Agilent Technologies) at the Iowa State University DNA facility.

Small RNA libraries tissue collection, construction and sequencing

_H. glycines_ TN19 and PA3 were propagated on the _susceptible Glycines max_ cultivar Williams 82 at Iowa State University. Nematode tissue was collected from each population in triplicate for the life stages egg, preparasitic J2, parasitic J2, J3, J4 and adult, using standard nematological methods (de Boer et al., 1999). The tissue was ground on liquid nitrogen. After RNA extraction, as listed above, 250ng of high quality total RNA from each sample was used as starting material in the NEBNext Multiplex Small RNA Library Prep Set for Illumina (New England Biolabs). Size selection for the miRNA peak ≈140bp was performed using a Novex 6% DNA gel (Thermo Scientific, Waltham, MA, USA) and SYBR Gold Nucleic Acid Gel Stain (Life Technologies). Sequencing was carried out by the DNA facility at Iowa State University using a HiSeq2500 (Illumina) on 50-cycle rapid run mode.

Small RNA Library Analysis

The raw reads were trimmed using Trimmomatic (Bolger et al., 2014). _H. glycines_ miRNA predictions were performed on the trimmed reads using miRDeep2 (Friedlander et al., 2012) and _H. glycines_ genome version 1 (V1) as generated by the ISU SCN Genome Group. Host target transcripts were predicted for all putative _H. glycines_ miRNAs with an average ≤ 10 reads for any life stage in at least one nematode population. miRNA target predictions were performed using psRNAtarget with the
following parameters: maximum expectation ≤ 3.0; G. max (soybean), unigene, DFCI Gene Index (GMGI, version 16).

Results

Secretion induction of H. glycines using chemical stimulants

H. glycines secretions were stimulated using varying concentrations of the serotonin receptor agonist Quipazine maleate salt. In H. glycines samples incubated with 0.25mg/ml of Quipazine maleate salt, 40-50% of nematodes showed head probing and stylet thrusting and had visible stylet secretions (Fig. 1A & 2). In samples containing only Coomassie blue, minimal head probing and stylet thrusting was

![Image A](image1.png)  ![Image B](image2.png)

Figure 1. Stylet secretions are induced in H. glycines when incubated with Quipazine maleate salt. Nematodes incubated in aqueous Coomassie for 4 hours at room temperature in the presence (A) or absence (B) of Quipazine maleate salt. The images are representative of three independent experiments, each with > 200 nematodes.
observed along with little to no visibly stained secretions (Fig. 1B & 2). The lowest concentration tested (0.1mg/ml), resulted in a low rate of head probing, stylet thrusting and reduced visibility of stylet secretions as compared to the 0.25mg/ml sample (Fig. 2). In higher Quipazine maleate salt concentrations (>0.25mg/ml), the head probing and stylet thrusting rates were comparable to the 0.25mg/ml sample (Fig. 2). However, it was also observed that exposure to higher levels of Quipazine maleate salt resulted in sporadic nematode movements in approximately 25%, potentially suggesting a high level of stress. Deviation from normal nematode movements suggests that the chemical had a large degree of off-target effects when present in high doses.

---

**Figure 2.** 0.25mg/ml Quipazine maleate salt is the optimum concentration to induce *H. glycines* stylet secretions with minimal off targeting effects. The percentage of *H. glycines* displaying head probing and stylet thrusting after 4 hours of treatment with varying concentrations of Quipazine maleate salt and staining with Coomassie Brilliant Blue G-250. The means and SE± are reflective of three independent experiments each with n=100, *P<0.05.*
Purification of Exosomes and RNA from *M. incognita* and *H. glycines* secretions

Induced and uninduced secretions from *M. incognita* and *H. glycines* were filtered alongside a nematode-free control and a cell serum sample containing exosomes. ExoQuick was used to purify exosomes from all of the filtered samples. As expected, microscopic analysis of the cell serum sample contained a large volume of exosomes (Fig. 3C). Surprisingly, both the induced and uninduced *M. incognita* and *H. glycines* secretions did not yield a high volume of exosomes (Fig. 3A and 3B). There was some evidence of exosome-like structures within the *M. incognita* secretions in a

---

**Figure 3.** Exosome purified *M. incognita* secretions reveal the low frequency and inconsistent appearance of spherical vesicles. Transmission electron microscopy of induced *M. incognita* (A), uninduced *M. incognita* (B), cell serum from an exosome positive cell line (C) and a no nematode buffer control (D). Samples of sporadic vesicle like structures in *M. incognita* induced secretion samples (E-H). The images are representative of three independent experiments, each with > 200 nematodes.
far lower volume than that observed in the positive control (Fig. 3E-H). In order to test if the lack of exosome-like structures was due to the method utilized, an identical set of *M. incognita* and *H. glycines* secretions, along with their relevant controls, were processed using the ultracentrifugation method. Consistent with the ExoQuick method, ultracentrifugation also produced a far lower volume of exosome-like structures in nematode samples as compared to the positive cell serum (data not shown).

---

**Figure 4. RNA isolation on exosome purified samples shows no detectable RNA in *M. incognita* secretions.** Small RNA Bioanalyzer profiles of induced *M. incognita* (A), uninduced *M. incognita* (B), cell serum from an exosome positive cell line (C) and a no nematode buffer control (D). The results are representative of three independent experiments.
RNA isolation was attempted on *M. incognita* samples, which were purified using the ExoQuick method. There was no detectable RNA in the induced or uninduced *M. incognita* samples or the negative control (Fig. 4A, 4B and 4D). A heterogeneous population of RNA starting at 16nt was observed in the cell serum sample (Fig. 4C).

**H. glycines** small RNA Libraries

The strategies used thus far in this study have focused on trying to isolate exosomes and small RNAs after secretion from the nematode. As a different approach, small RNA libraries were generated using whole nematode tissue from two different populations, TN19 (Hg type 1-7) and PA3 (Hg type 0), to identify miRNAs with the

---

![Graph showing miRNA read counts throughout the H. glycines life cycle.](image)

**Figure 5.** miRNA read counts throughout the *H. glycines* life cycle. miRNA predictions were performed using miRDeep2 for two *H. glycines* populations, TN19 (left) and PA3 (right). Read counts are reflective of three biological replicates for each life stage and population. Graphs represent the total read counts across all life stages of both populations for the 21 miRNA predictions were the average read count was ≤ 10 in at least one life stage of either population.
potential to target host genes. Nematode tissue was collected at six different time points to reflect the small RNA populations during the egg, preparasitic J2 (migration outside the root tissue), parasitic J2 (migration through the root tissue and feeding site establishment), J3, J4 and adult life stages.

After application of miRDeep2, using the *H. glycines* genome as a reference, a total of twenty-one miRNAs were predicted, ten of which have reads in all life stages of both populations (Fig. 5). Alignment of the twenty-one predicted miRNAs yielded a total of nineteen unique miRNAs, indicating that two of the predicted miRNAs mapped to two different locations within the *H. glycines* genome (Fig. 6).

---

**Figure 6. sequence alignment of predicted miRNAs.** A multiple sequence alignment of predicted miRNAs was performed using T-coffee with default settings, (-) indicates a sequence.
In order to predict host targets in *Glycines max*, the nineteen unique putative *H. glycines* miRNAs, were used as input sequences for psRNATarget against the *G. max* transcriptome. psRNATarget identified a total of 200 host targets, when using a less conservative Expectation score (E ≤ 3.0), with each putative miRNA having at least one predicted host target (Table 1). To prioritize candidates for future functional characterization a more conservative E ≤ 2.0, was applied to the dataset, reducing the number to fifteen predicted *G. max* targets and seven *H. glycines* putative miRNAs (Table 2).

<table>
<thead>
<tr>
<th><em>H. glycines</em> miRNA</th>
<th>Total number of predicted targets in <em>Glycine max</em></th>
<th>E ≤ 2.0</th>
<th>E ≤ 3.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>126500-126600</td>
<td>9</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>142800-143500</td>
<td>6</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>143000-143100</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>164400-165100</td>
<td>4</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>169200-169300</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>184400-184400</td>
<td>17</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>187200-187300</td>
<td>9</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>191000-191100</td>
<td>4</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>193300-193400</td>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>197200-197300</td>
<td>30</td>
<td>2</td>
<td>28</td>
</tr>
<tr>
<td>204200-204300</td>
<td>20</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>246600-246700</td>
<td>30</td>
<td>2</td>
<td>28</td>
</tr>
<tr>
<td>282300-282400</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>366900-367000</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>424800-424900</td>
<td>38</td>
<td>4</td>
<td>34</td>
</tr>
<tr>
<td>526700-535200</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>593600-593700</td>
<td>4</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>702600-703200</td>
<td>17</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>920400-921200</td>
<td>5</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td><strong>200</strong></td>
<td><strong>15</strong></td>
<td></td>
<td><strong>185</strong></td>
</tr>
</tbody>
</table>
Discussion

A collection of nineteen previously published exosomal proteomic studies from various cell types highlighted a conserved set of exosomal proteins (Mathivanan et al., 2010). The conserved set included heat shock proteins, actin, enolase and protein disulfide isomerase all of which were found in the top ten most abundant peptides secreted by *M. incognita* (Bellafiore et al., 2008; Mathivanan et al., 2010). Here we used
a similar induction method to analyze *M. incognita* secretions for the presence of exosomes. The secretions of another sedentary PPN, *H. glycines*, were also induced and collected using a novel serotonin receptor agonist Quipazine Maleate Salt. This novel chemical displays similar rates of head probing and secretion induction to those published using DMT, a chemical that has since been discontinued (Goverse et al., 1994).

Exosome isolation of induced PPN secretions revealed the presence of a small number of structures with a similar size and shape to previously identified pathogen-derived exosomes (Twu et al., 2013; Zamanian et al., 2015). These vesicles were not present in the uninduced and nematode free samples, suggesting that they are part of the PPN secretions. The low volume of vesicles observed in the induced PPN samples was not due to technical error in the exosome isolation step as the cell serum displays a large volume of exosomes. It should be noted that the cell serum contained millions of cells prior to exosome isolation while the PPN secretions were collected from around 1 million nematodes. It is possible that the amount of nematode starting material is inadequate for exosome detection using these methods. Increasing the amount of nematode material poses a significant challenge as only 1 in 10 PPN will hatch using current lab techniques and only half of these are likely to produce secretions using chemical induction.

Sedentary endoparasitic PPN have six life stages and their infection involves two distinct phases; migration inside the root tissue and initiation of a permanent feeding structure (Goverse and Smant, 2014). The identification of a low volume of exosomes from the freshly hatched PPN tested in this study could suggest that PPN
do not secrete exosomes during migration towards the root tissue. It is already known that PPN infection is a dynamic process as proteins secreted by PPN, referred to as effectors, display differing expression profiles across life stages (Hewezi et al., 2010; Xie et al., 2016). Differences in the abundance and appearance of exosome-like vesicles between the larval and adult stages of the human filarial nematode *B. malayi* also support the idea that all nematode stages may not use exosomes (Zamanian et al., 2015). In the case of PPN, exosomes may only be utilized in the establishment of a more intimate relationship with specific cells during feeding site initiation. To test if exosomes are crucial for feeding site initiation, exosome purification would have to be performed on the later stages of PPN, which can only be harvested from root tissue. Separation of PPN from host root tissue would have technical challenges as far as collecting adequate amounts of material.

The discovery of twenty-one putative *H. glycines* miRNAs, followed by subsequent prediction of their transcript targets within the host transcriptome *G. max* has revealed the potential for PPN-derived miRNAs to directly alter transcript levels within the host plant. After applying a conservative cutoff of $E \leq 2.0$ to the transcript target predictions, there were fifteen high priority candidates for future functional characterization. It is interesting to note that two of the PPN-derived miRNAs are each predicted to target four host transcripts. This would suggest that *H. glycines* is capable of secreting a small set of miRNAs to simultaneously target a large number of host transcripts. Another interesting finding is the commonalities between perceived functions of the predicted host transcript targets and previously identified PPN-infection strategies. For instance, one of the putative *G. max* transcript targets is most
similar to tonoplast intrinsic aquaporin 4-1 (TIP4-1). Aquaporins are differentially regulated within nematode feeding sites and the *M. incognita* effector Mi8D05 has been shown to physically interact with a TIP (Klink et al., 2005; Opperman et al., 1994; Xue et al., 2013). Future functional confirmation of *H. glycines* miRNAs within host cells will represent a significant breakthrough in our understanding of how PPN are able to induce such extensive alterations in host gene expression.

**References**


de Boer, J. M., Y. T. Yan, X. H. Wang, G. Smant, R. S. Hussey, E. L. Davis, and T. J. Baum, 1999, Developmental expression of secretory beta-1,4-endoglucanases in the subventral


Klink, V. P., N. Alkharouf, M. MacDonald, and B. Matthews, 2005, Laser capture microdissection (LCM) and expression analyses of Glycine max (soybean) syncytium
containing root regions formed by the plant pathogen Heterodera glycines (soybean cyst nematode): Plant Molecular Biology, v. 59, p. 965-979.


CHAPTER 4. HETERODERA GLYCINES UTILIZES HYPERVERSIBLE SPICED LEADERS TO PROMISCUOUSLY SL TRANS-SPLICE A LARGE PORTION OF TRANSCRIPTS

Modified from a manuscript to be submitted for publication to the journal RNA

Stacey N. Barnes, Rick E. Masonbrink, Thomas R. Maier, Anoop S. Sindhu,
Andrew J. Severin, Thomas J. Baum

Abstract

Spliced leader trans-splicing (SLTS) is a prevalent mechanism involved in the maturation of pre-mRNAs that is found across multiple phyla, but is particularly prevalent in Nematoda. The role of spliced leaders (SL) within the cell is unclear and an accurate assessment of SL occurrence within an organism is possible only after extensive sequencing data are available, which is not currently the case for many nematode species. SL discovery is further compromised by an absence of SL sequences from high-throughput sequencing due to incomplete sequencing of the 5’ ends of transcripts, known as 5’ bias, during RNA-seq library preparation. Existing datasets and novel methodology were used to identify both conserved SLs and unique H. glycines hypervarsible SLs. Twenty-one SL sequences were found in H. glycines on >2,000 transcripts, nearly half of which were trans-spliced to multiple SLs. Mapping of the SL trans-spliced transcripts to the H. glycines genome indicated several large clusters of SL trans-spliced genes, and functional analysis of the SL trans-spliced
transcripts shows involvement in a wide range of biological processes, including ‘Embryo development ending or egg hatching’ and ‘Nematode larval development’. The discovery of such a large and partially unique set of SLs within *H. glycines* and their involvement in key biological processes suggests that SLs play a key role in *H. glycines* gene expression.

**Introduction**

Pre-mRNA splicing is a vital mechanism associated with the expression and regulation of eukaryotic genes. The most widely deployed splicing mechanism is cis-splicing, which enables the removal of intron sequences from within a single mRNA molecule. Trans-splicing is less widespread and results in the fusion of RNA molecules that are transcribed from different genomic loci. The most prevalent form of trans-splicing involves the addition of a short nucleotide spliced leader (SL) sequence to the 5’ end of mRNA transcripts, referred to as spliced leader trans-splicing (SLTS). SLTS has evolved independently in a diverse set of phyla including: *Nematoda, Platyhelminthes, Trypanosoma, Cnidaria, Rotifera, Chordata, Arthropoda* and *Dinoflagellata* (Douris et al., 2010; Ganot et al., 2004; Krause and Hirsh, 1987; Pouchkina-Stantcheva and Tunnacliffe, 2005; Rajkovic et al., 1990; Stover and Steele, 2001; Vandenberghe et al., 2001; Zhang et al., 2007).

SLs originate from SL RNA genes, which are comprised of two parts divided by a donor splice site: the 5’ exon-like SL region and a 3’ intron-like region (Bruzik et al., 1988; Hannon et al., 1992; Sharp, 1987). The SL RNA maintains a conserved
secondary structure comprised of hairpins and a single-stranded Sm binding site (5’-purine-AU₄₆G-purine-3) to ensure that the SL RNA can interact with proteins that are required for SLTS (Krause and Hirsh, 1987; Riedel et al., 1987; Thomas et al., 1990).

It is evident that SLTS plays a role in resolving polycistronic mRNAs in Caenorhabditis elegans, acting as a prerequisite for subsequent translation (Spieth et al., 1993). In C. elegans, approximately 70% of transcripts are trans-spliced to one of two 22nt SLs: SL1 and SL2 (Allen et al., 2011; Huang and Hirsh, 1989; Krause and Hirsh, 1987; Zorio et al., 1994). However, operon resolution is not the sole function of SLTS in C. elegans as only 17% of C. elegans transcripts originate from operons (Allen et al., 2011). Some hypotheses suggest that SLTS is involved in various forms of translational regulation including replacement of deleterious sequences in the 5’-untranslated region, addition of translational motifs from within the SL sequence, or the replacement of a transcript’s 5’-monomethylated cap with a hypermodified 5’-cap structure (Danks et al., 2015; Hastings, 2005; Lall et al., 2004; Maroney et al., 1995).

All nematode species studied to date show some evidence of utilizing SL1, SL2, or close variants, with the exception of Trichinella spiralis which uses its own non-canonical spliced leaders (Cotton et al., 2014; Goyal et al., 2005; Harrison et al., 2010; Mitreva et al., 2004; Nilsen et al., 1989; Pettitt et al., 2008; Ray et al., 1994; Takacs et al., 1988). Interestingly, sequence analysis of the potato cyst nematode Globodera rostochiensis and G. pallida, identified multiple hypervariable SL sequences in addition to SL1 and SL2 (Bers, 2008; Cotton et al., 2014). The diversity of SL sequences found in Globodera spp. and the dearth of information regarding their functionality in the cell highlights the need to expand nematode genomic and transcriptomic investigations to
elucidate the role of SLs in nematode functionality, viability and parasitism. The genome and multi-stage transcriptome of the soybean cyst nematode, *H. glycines*, have recently been made available, making *H. glycines* a strong candidate for SL analysis. SL1 has previously been identified in *H. glycines* genomic contigs (Fosu-Nyarko et al., 2016) and the SL1 sequence has been used to successfully generate *H. glycines* cDNA libraries (LIBEST_005577). *H. glycines* is also a closely related species to *Globodera spp.* meaning that it is highly likely to contain SLs.

This study utilized the *H. glycines* genome and transcriptome to extensively characterize SLs and their usage in *H. glycines*. Variations in the 5’-end of a previously sequenced *H. glycines* transcript led to the initial discovery of a novel SL. Through subsequent bioinformatics approaches utilizing both *H. glycines* genomic and transcriptomic data, this report shows that *H. glycines* possesses unique hypervariable SLs on a large number of transcripts. Functional analysis of the SL trans-spliced transcripts revealed involvement in a variety of biological processes.

**Materials and Methods**

**Identification and structure prediction of HgSL3 RNA**

The spliced leader-containing transcript MH119144 and its 22nt spliced leader sequence were queried against the completed and annotated *H. glycines* genome Version 1 (V1) with BLASTn (E-value 1.0E-3). The sequences which aligned to the first 22nt of MH119144 and the 98 nucleotides 3’ of each hit were extracted using Samtools. Secondary structure was predicted using RNAfold V2.1.9 with unpaired
bases participating in at most one dangling end.

**DNA extraction and amplification**

To confirm the functionality of putative SL on transcript MH119144 OP50 *H. glycines* nematodes were propagated on Williams 82 soybean in the greenhouse. To isolate mixed-stage nematodes, root tissue was macerated with a blender, sieved and separated with a sucrose gradient (de Boer et al., 1996). The nematodes were ground in liquid nitrogen and total RNA was extracted using a RNeasy Mini Kit (Qiagen). One μg of total RNA was treated with DNase I (Thermo Fisher Scientific) and cDNA was synthesized using qScript cDNA SuperMix (Quantabio, Beverly, MA, USA). Genomic DNA was also extracted from ground nematode tissue using QIAamp DNA Mini Kit (Qiagen). RT-PCR was performed on a Bio-Rad S1000TM thermal cycler with reactions containing 1X PCR buffer, 1.5mM MgCl2, 0.2mM dNTP, and 1 unit of Taq DNA Polymerase (ThermoFisher Scientific). Thermocycler conditions were: 94°C for 3min, 35 cycles of 95°C for 45s, 55°C for 30s and 72°C for 1min, followed by 10min at 72°C.

**Genome-wide SL assessment**

All *Globodera rostochiensis* SLs (Cotton et al., 2014) and all SL sequences present in the *C. elegans* genome assembly WBcel235 (PRJNA13758) were used as queries in BLASTn searches against the *H. glycines* genome V1 (E-value 1.0E-3). The genomic sequences including the 98 nucleotides 3’ of all exact matches were extracted with Samtools. All extracted sequences were analyzed for a downstream Sm motif (5’-purine-AU₄₋₆G-purine-3’) (Thomas et al., 1988).
A *H. glycines* spliced leader RNA gene was queried against the *H. glycines* genome (E-value 1.0E-3; ≥80nt alignment length) with BLASTn. Sm motif searches and RNA folding predictions were performed as described earlier.

**Identification of SL trans-spliced transcripts**

The newly discovered *H. glycines* SLs and all other SLs that were queried against the *H. glycines* genome were used as query sequences for transcript analysis. The dust parameter was applied to all SL BLASTn queries to the *H. glycines* NCBI EST database and *H. glycines* trinity transcriptome V1. BLAST hits were filtered to contain results in the forward orientation, within the first thirteen nucleotides of the transcript, and with an alignment length minimum of ten. ESTs were mapped to trinity transcripts with GMAP to consolidate all sequence comparisons into Trinity transcripts.

**Read Analysis**

SLs were truncated from the 5’ end, leaving only eleven nucleotides at the 3’ end. The truncated SLs were queried with BLASTn without dust against the paired-end reads that were used to generate the *H. glycines* trinity transcriptome V1. The subject start position for all read hits was extracted and graphed using GraphPad Prism 4. High stringency was added to the BLAST output by applying the same filter described earlier: ($10 > $9 && $9 < 13 && $4>10) or ($9 > $10 && $9 > 88 && $4>10). Filtered reads were subsequently queried with BLASTn (E-value etc) against the *H. glycines* Trinity transcriptome V1.
Spliced Leader Transcript Clustering

Trinity transcripts were mapped to the genome using Gmap with default parameters. Bedtools intersect identified exonic-overlap with transcripts and genes. Genes were clustered by location using custom bash scripts.

Functional Analysis

Functional annotation was performed using Blast2go version 4.1. All sequences were searched against the NCBI non-redundant protein database using Blastx with an E-value cutoff of 1.0E-5. Interpro scan was preformed using all default selected applications. The sequences were annotated using an annotation cutoff of 55 and a GO weight of 5. GO enrichment for trans-spliced genes was performed using Ontologizer with the published functional gff for the H. glycines genome.

Trans-splicing in Effector and Repetitive Genes

Bedtools intersect and custom bash scripts were used to identify trans-spliced repetitive genes from a published Repeatmodeler (Quinlan and Hall, 2010). Effector genes were mapped to the genome using Gmap and bedtools and custom bash scripts were used to identify effectors subject to trans-splicing.
Results

Discovery of a novel spliced leader in *H. glycines*

While analyzing available *H. glycines* transcriptome data, two transcripts (AY160225 & MH119144) were identified as chorismate mutase, a gene that is involved in parasitism in multiple plant-parasitic nematodes (Bekal et al., 2003; Doyle and Lambert, 2003; Gao et al., 2003; Jones et al., 2003; Lambert et al., 1999; Vanholme et al., 2009). AY160225 was initially cloned and sequenced after amplification using a SL1 forward primer and a gene-specific reverse primer located in a well-conserved region (Bekal et al., 2003). MH119144 was amplified using SMARTer technologies to sequence the complete 5’ end. Sequence alignment of SL1 and the 5’ end of MH119144 revealed that the two 5’ ends were divergent. Alignment of the 5’ end of MH119144 to the well-documented *C. elegans* SL1 sequence led to the hypothesis that the MH119144 transcript has a novel SL sequence (Fig. 1A).

To further investigate the putative MH119144 SL sequence, the entire transcript was queried against the *H. glycines* genome with BLASTn. All but the first fifteen nucleotides of MH119144 mapped to scaffold_282, supporting the presence of a novel SL (Table S1, Appendix B). The putative twenty-two nucleotide SL at the 5’ end of MH119144 had four exact hits in the *H. glycines* genome all of which mapped within a 2.5Kb region on scaffold_362 (Table S2, Appendix B). In order for a SL to be functional, transcription must create a distinct non-coding hairpin SL RNA structure with a single-stranded Sm motif (Krause and Hirsh, 1987; Riedel et al., 1987; Thomas et al., 1990). To identify the presence of these features, the ninety-eight nucleotides downstream of
the four SL hits was extracted. Alignment of the four sequences showed 99% similarity to each other and the typical secondary structure of functional SL RNAs (Fig. 1B & S1, Appendix B).

The functionality of the putative SL was tested further using RT-PCR to search for the putative SL chorismate mutase sequence in *H. glycines* gDNA and cDNA (Fig. 1C). Using the putative SL sequence as a forward primer and a gene-specific reverse primer, a visible band was produced when using a cDNA template, but not gDNA (Fig. 1C). Genic structure predictions performed on chorismate mutase indicate that the
absence of a band within the gDNA reaction is not due to the SL primer being located on intron/exon border. Furthermore, a control PCR amplification with cDNA and gDNA templates was performed using a gene-specific primer pair to verify the presence of the chorismate mutase gene in both DNA samples (Fig. 1C). Collectively, mapping of the putative SL and the remainder of the transcript to separate locations within the genome, the similarity of the putative SL RNA sequence to known SL RNAs and the absence of a SL chorismate mutase PCR product when using gDNA all support the functionality of the SL, which will therefore be referred to as *Heterodera glycines* spliced leader 3 (HgSL3).

**The *H. glycines* genome contains multiple novel SL sequences**

To investigate the existence of previously identified SL sequences in *H. glycines*, all known *C. elegans* and *Globodera spp.* SLs were mapped to the *H. glycines* genome. SL1 mapped to 180 loci in the *H. glycines* genome, twenty-two sequences of which were located within close proximity to the essential Sm motif (Table S3, Appendix B) (Riedel et al., 1987). However, while *Globodera spp.* SL1b mapped to the genome, it lacked a proximal Sm motif. All other *Globodera spp.* SLs were absent from the genome (Table S3, Appendix B).

The reduced conservation of existing SLs within the *H. glycines* genome suggests that *H. glycines* may possess a unique set of novel SLs alongside SL1 and HgSL3. BLAST was used to search for novel SL genes in the *H. glycines* genome, using the HgSL3 RNA gene as a query (Table S4, Appendix B). A total of sixty-nine sequences were identified that also contained Sm-binding sites. After analyzing their
secondary structures, twenty sequences folded to include a single-stranded Sm-binding site flanked by hairpins, including four copies of the HgSL3 RNA gene (Table S4, Appendix B). Alignment of the first twenty-two nucleotides of the putative HgSL RNA sequences yielded the HgSL3 sequence and ten putative HgSLs, subsequently numbered HgSL4-13 (Fig. 2).

<table>
<thead>
<tr>
<th>HgSL3</th>
<th>GCAATTAATTTGTCACAATTAAG</th>
</tr>
</thead>
<tbody>
<tr>
<td>HgSL4</td>
<td>GACAAGAATTTGTCAGTTAAG</td>
</tr>
<tr>
<td>HgSL5</td>
<td>TGCAATTAATTTGTCACAATTAAG</td>
</tr>
<tr>
<td>HgSL6</td>
<td>GACAATTAATTTGTCACAATTAAG</td>
</tr>
<tr>
<td>HgSL7</td>
<td>GACAAGAATTTGTCACAATTAAG</td>
</tr>
<tr>
<td>HgSL8</td>
<td>TGGCTCTTCTGTAAG</td>
</tr>
<tr>
<td>HgSL9</td>
<td>TGGCTCTTCTTAGCTCTTAAG</td>
</tr>
<tr>
<td>HgSL10</td>
<td>GCTTTGCTTTGTGCTCTTAAG</td>
</tr>
<tr>
<td>HgSL11</td>
<td>ATGGGTCTTTCTGTGCTCTTAAG</td>
</tr>
<tr>
<td>HgSL12</td>
<td>CATTGGCA-TTA-TTGT--CCTTAAG</td>
</tr>
<tr>
<td>HgSL13</td>
<td>GTGGCAATTATTTGTCACAATTAAG</td>
</tr>
</tbody>
</table>

**Figure 2 - Sequence alignment of novel hypervariable *H. glycines* SLs.** Multiple sequence alignment of *H. glycines* HgSL3 RNA blast results with SL RNA-like secondary structural predictions. The alignment was performed using T-coffee with default settings. (*) indicates a consensus nucleotide position and (-) denotes a sequence gap. Pink, yellow and green shading indicates good, average and bad nucleotide matches respectively.

SL sequences from *H. glycines*, *C. elegans* and *G. rostochiensis* are promiscuously present on *H. glycines* transcripts

To assess SL usage, *H. glycines* transcripts were analyzed for the presence of SLs at their 5' ends. In order to keep the analyses inclusive, all *C. elegans* and *Globodera spp.* SLs were utilized alongside the novel *H. glycines* SLs, irrespective of
whether the SLRNA was identified in the *H. glycines* genome. To carry out the transcript analysis, the SLs were first truncated to contain only the last (3') 11nt, yielding a total of twenty-six distinct SL queries (7 from *C. elegans*, 5 from *H. glycines*, and 14 from *Globodera spp*.). The use of truncated SLs has previously been utilized in *G. pallida* when searching for SL-containing reads, as a low availability of 5' ends has previously been hypothesized as a limitation to SL identification (Cotton et al., 2014; Pettitt et al., 2008).
The 11nt truncated SL sequences were utilized in a three-part blast analysis using different databases to identify SL trans-spliced transcripts. The truncated SL sequences were first used as BLAST queries to ESTs, secondly to a *H. glycines* Trinity transcriptome, and finally to illumina reads that were then mapped to transcripts. In the final case, the novel read-to-transcript approach can circumvent the issue of assembled transcripts lacking a complete 5’ end, which is a common artifact of library preparation (Lahens et al., 2014). BLAST searches to ESTs and transcripts yielded 187 and 2,076 unique SL trans-spliced transcripts, respectively (Table 1). A small number

![Venn diagram](image)
of the 2,076 transcripts received more than one SL sequence, bringing the total to 2,215 SL-transcript combinations (Table 1 & Fig. 3). Finally, the read-to-transcript approach identified 85,876 of ~11.4 million reads that had a terminal SL, the legitimacy of which is supported by SL BLAST hits preferentially locating to the 5’ read ends (Fig. 4). The reads that possessed a terminal SL were subsequently mapped to the

![Graph](image)

**Figure 4 - HgSLs are preferentially located at the 5’ end of *H. glycines* reads.** All truncated HgSLs were queried with BLASTn against the *H. glycines* raw reads, which were also used for the Trinity assembly. The read nucleotide start positions were plotted to show the strong preference for HgSLs to be located at the 5’ end of the read.

Trinity transcripts using BLAST to identify 1,635 unique SL trans-spliced transcripts. Again, a portion of the transcripts received more than one SL, resulting in 6,350 SL-transcript combinations (Table 1 & Fig. 3). Collectively, these analyses resulted in 2,532 unique SL trans-spliced transcripts and 21 functional SLs (Table 1 & Fig. 3). Interestingly, when combining all three analyses, 45.5% of the 2,532 SL trans-spliced
transcripts were spliced to two or more SLs with 6.8% of transcripts being trans-spliced by five or more different SLs (Fig. 5).

**Figure 5** - **SL trans-spliced transcripts are promiscuously trans-spliced.** The number of different truncated SL sequences identified at the 5' end of all unique SL trans-spliced transcripts.

**Genomic features of genes that possess spliced leaders**

To identify genomic features that may be associated with SL trans-spliced transcripts, the transcripts were mapped to the *H. glycines* genome using Gmap. Exonic overlap between *H. glycines* genes and SL trans-spliced transcripts accounted for ~1/3 of the genes in the genome (12,060/29,959). To assess the positioning of SL trans-spliced genes within the *H. glycines* genome, the genome was partitioned into 50kb stretches. Analysis of the 50kb genomic segments showed that SL trans-spliced
genes were dispersed throughout the genome. However, clustering of SL trans-spliced genes was also evident, as forty of the 50kb stretches showed fourteen or more consecutively arranged SL trans-spliced genes (Table S5, Appendix B).

**Functional analysis of SL trans-spliced transcripts reveals involvement in a variety of biological processes**

In order to gain functional insight into the role of SL trans-splicing in *H. glycines*, the SL trans-spliced transcripts were annotated with Blast2go (Conesa et al., 2005). Over half (52%) of the annotated transcripts were involved in metabolic and developmental processes (Fig. 6), with the top two biological processes involved in ‘Embryo development ending or egg hatching’ and ‘Nematode larval development’ (Fig. S2, Appendix B). A complementary GO enrichment analysis was performed on the corresponding genomic genes, revealing a similar profile of functions involved in metabolic processes (Table S6, Appendix B).

It is interesting to note that of the 12,060 SL trans-spliced genes in the genome, ~1/4 (3,218) were associated with a genomic repeat suggesting that repetitive elements may also be subject to SLTS. Nine repeats comprised 22% of the 3,218 repeat-associated trans-spliced transcripts, with the most abundant repeat being an unknown element comprising 6.7% of the total. A LINE/CR1 retrotransposon was the most abundant annotated SLTS repetitive transcript at 4.3% of the total, suggesting that transposon-derived transcripts are also subject to SL trans-splicing (Table S7, Appendix B).
To gain insight into how spliced leaders could be involved in parasitism, we searched for exon-exon overlap between SL trans-spliced transcripts and effector genes in the genome. Effector genes produce proteins that are secreted by *H. glycines* throughout parasitism to play a major role in altering host cell structure and function (Reviewed by Davis et al., 2008; Hewezi and Baum, 2013; Juvale and Baum, 2018; Mitchum et al., 2013). Genes encoding SL trans-spliced transcripts overlapped with
twenty-nine of 121 effector genes in the genome. Eight of the twenty-eight multiple-copy effector genes in the genome have SL trans-spliced transcripts (Table S8, Appendix B). However, there appears to be no correlation between gene copy number and SL trans-splicing as 5 of 5 genes are trans-spliced for the effector 11A06, while 0 of 5 genes are trans-spliced for the effectors 4D06 and 32E03.

**Discussion**

This study identified and characterized SLs and their usage within the plant-parasitic nematode *H. glycines*. The recent availability of both the *H. glycines* genome and transcriptome has provided a great opportunity to extensively characterize SL use and function in a parasitic nematode.

This study was prompted by the discovery of HgSL3 at the 5’ end of a chorismate mutase cDNA, leading to the identification of a unique set of hypervariable HgSLs. Large sets of novel hypervariable SLs have previously been discovered in the plant-parasitic nematode *Globodera rostochiensis* and the animal-parasitic nematode *Trichinella spiralis* (Cotton et al., 2014; Pettitt et al., 2008). Interestingly, despite the high volume of SLs that have been discovered in these three species, genomic data suggests a low interspecies conservation of SLs. Given the parasitic nature of all three species, as well as the perceived link between SLs and translational regulation, it is possible that large hypervariation of SLs is a response to parasitism of different hosts. This study investigated a possible link between SLs and known parasitic molecules, referred to as effectors, but no correlations were found. Effector discovery in *H.*
glycines is still ongoing so it is possible that a connection between SL trans-splicing and effectors may be uncovered in the future (Juvale and Baum, 2018).

To identify H. glycines SL trans-spliced transcripts, the SLs were first truncated before being queried using BLAST against H. glycines sequences. The use of truncated SLs was previously utilized in G. pallida (Cotton et al., 2014). Before using truncated SLs in H. glycines, we first verified that this approach was necessary by using the full length SLs as query sequences against the H. glycines ESTs and Trinity transcriptome. Only 15 sequences, none of which were SL1, were identified across both databases when using full length SLs (data not shown). The absence of full length SL1 within the H. glycines databases supports the hypothesis that there is a lack of 5’ ends within the H. glycines datasets, as SL1 is documented to be a highly utilized SL in other nematodes within the H. glycines clade (Cotton et al., 2014; Mitreva et al., 2004). This study further illustrated the lack of complete 5’ ends within the H. glycines transcriptome by showing that the truncated SLs were predominantly located at the first nucleotide of the raw reads. This means that the first 11nt of the mature transcript, which contained the full length SL, are often underrepresented in the raw reads. To further complicate transcriptome assembly in SLTS organisms, this study revealed that 45.5% of SL trans-spliced transcripts receive more than 1 SL. The promiscuous nature of SLs on otherwise identical transcripts may cause high ambiguity in the assembly step, resulting in 5’ truncation of the transcript or assembly of a transcript that reflects only the most highly expressed SL-transcript version while discarding lower expressed SL-transcripts.
Analysis of the available *H. glycines* ESTs, Trinity transcriptome and raw reads used in this study concluded that HgSL3 is the most prevalent SL in *H. glycines*, with 30.9% of the SL trans-spliced transcripts being trans-spliced by HgSL3. The predominant use of a non-SL1 sequence in *H. glycines* contrasts with findings in *C. elegans* and *Ascaris suum*, as well as *G. pallida* where SL1 and SL1 variants were identified on >90% of the SL-containing *G. pallida* reads (Cotton et al., 2014).

*C. elegans* operon genes, which are resolved into monocistronic transcripts using SL trans-splicing, are upregulated during recovery from growth-arrested states (Spieth et al., 1993; Zaslaver et al., 2011). Operon arrangement is believed to be advantageous in *C. elegans* during times of limited resources as there are less promoters competing for transcriptional resources (Zaslaver et al., 2011). In the case of *H. glycines*, SL trans-spliced transcripts were found to be involved in ‘Embryo development ending or egg hatching’ and ‘Nematode larval development,’ suggesting that SL trans-splicing may also play a role in initiating developmental changes in *H. glycines*. Operon arrangement has not yet been defined in *H. glycines*, however the clustering of SL trans-spliced transcripts in the genome suggests the presence of operon-like structures.

In summary, *H. glycines* possesses a unique set of hypervariable SLs, which, alongside some previously known SLs, are promiscuously trans-spliced to the 5’ end of many *H. glycines* transcripts. A robust identification of SLs was possible through novel methodology and the availability of *H. glycines* genome and transcriptome sequences. As more data become available for *H. glycines* and other parasitic and
non-parasitic nematodes, the functional significance of SLTS may become more apparent and potentially lead to novel control measures.

References


CHAPTER 5. GENERAL CONCLUSIONS

Throughout this dissertation it is clearly described that *Heterodera glycines* parasitism involves a highly complex and intimate relationship between the nematodes and their host plants, which involves a significant alteration to many host processes. The mechanisms driving such dramatic alterations in the host are currently under investigation and aim to provide novel ways in which the host can be manipulated to favor crop resistance.

One key mechanism that has been linked to successful plant-parasitic nematodes (PPN) infection is the secretion of proteinaceous effectors by *H. glycines* into various host subcellular compartments. In the last twenty years much progress has been made relating to the identification of effector proteins. Subsequent in depth functional characterization of a subset of effector proteins has helped to uncover key host processes that are targeted by *H. glycines* including the breakdown of cell walls, hormone manipulation and suppression of defenses. Interestingly simultaneous effector identification and functional characterization within other PPN species is uncovering both commonalities and unique infection strategies deployed by PPN.

To date, there has been little research performed pertaining to how effectors function once inside the host plant nucleus. This knowledge gap provided adequate motivation for an in-depth functional analysis of the host-nuclear localized effector GLAND4, listed in chapter two. The study identified that GLAND4, once inside the plant nucleus, is the first plant-parasitic nematode (PPN) effector that is capable of binding
to host DNA. The host DNA region bound by GLAND4 is flanked by two lipid transfer protein (LTP) genes, both of which are downregulated in the presence of the effector. Downregulation of the LTPs added further novelty to the study by recognizing GLAND4 as the first plant-pathogenic DNA-binding repressor with a known in vivo host target. The two LTPs were already known to be differently regulated in response to various pathogens, a finding that was further corroborated during the course of this GLAND4 study. The involvement of these particular LTPs in multiple plant-pathogen interactions makes protection of LTP expression levels and host targeting of GLAND4 attractive solutions for enhanced host resistance in the future.

The knowledge that multiple plant pathogens often utilize similar infection mechanisms, such as DNA-binding, to promote successful host infection provided a basis by which to search PPN secretions for exosomes and small RNAs. The secretion of exosomes and small RNAs is a well-documented strategy in many pathogens including multiple animal-parasitic nematodes. Exosome isolation performed on PPN secretions yielded a low number of inconsistent vesicular structures of similar size and shape to those previously documented in animal-parasitic nematodes. The recovery of such structures within PPN secretions provides a solid groundwork for future studies that may see an increased yield with the inclusion of more life stages and optimized protocols for minute quantities of starting material.

Investigations into the use of exosomes as a bone fide parasitic strategy would also be greatly enhanced by microscopic visualization of PPN during their infection stages. Many of the existing PPN microscopy images, which enhanced our understanding of the anatomical features of PPN, are now several decades old, during
which time there have been many advances in both the preparation of tissue and tissue imaging. Observations at the time acknowledged that structures within the gland cells, where it is hypothesized that the parasitic exosomes are derived from, differed in appearance but the fixation methods were not optimized for the analysis of subpopulations with the gland cells.

The low quantity of exosome recovery proved problematic with regards to subsequent isolation of secreted small RNAs. Instead, through the isolation of small RNAs from all life stages of two different *H. glycines* populations it was bioinformatically predicted that *H. glycines* express miRNAs that are capable of targeting host transcripts. Observing the functions of predicted host transcript targets of *H. glycines* miRNAs has revealed potential overlap with pathways that have previously been identified as targets for modification by PPN through the secretion of effector proteins. This suggests that *H. glycines*, and likely other PPN, can ensure robust targeting of host pathways by using both post-transcriptional and post-translational targeting strategies. Future studies to biologically confirm these predictions will open up a new avenue of research into plant-nematode interactions.

Based on the results in chapter three of this dissertation, small RNA sequencing of the two different nematode populations, one virulent and one avirulent both grown on susceptible plants, yielded high uniformity in identified miRNAs between the two populations. Assessing the expression of a subset of these miRNAs across a multitude of nematode populations will help to determine how influential miRNA secretion is in determining parasitic success. Conducting similar large scale small RNA sequencing
analyses using both susceptible and resistant plants will also shed light on the influence that miRNA secretion plays in parasitism.

In recent decades, the development and constant improvement of high-throughput sequencing technologies has created a greater availability of genomic and transcriptomic data for a wide array of organisms, including *H. glycines*. The spliced leader (SL) project, described in chapter four of this dissertation, utilized the new *H. glycines* sequence data to enhance our understanding of gene regulation and signaling within *H. glycines*. Unique methodology was applied to the genome and transcriptome to discover novel hypervariable *H. glycines*-specific SLs and one of the largest SL trans-spliced transcript populations known in *Nematoda*. Analysis of the SL trans-spliced transcripts revealed that the majority of transcripts were trans-spliced to either SL1, which is widely conserved within the nematode phylum, or HSL3, which is novel to *H. glycines*. Interestingly, the SLs also demonstrated high promiscuity, meaning that a particular transcript was often trans-spliced by multiple different SLs. The high volume of SL trans-spliced transcripts, the apparent effort to diversify SL sequences and the promiscuity of SLs all suggest that SL trans-splicing is a necessary mechanism involved in *H. glycines* signaling. A greater breadth of SL studies in nematodes from numerous clades within the phylum will help to assess the importance of SL trans-splicing for nematode viability. Studies that can encompass both parasitic and non-parasitic species may uncover differences in SL signaling, which may ultimately lead to a novel resistance mechanism that can simultaneously interfere with a large number of *H. glycines* transcripts.
In summary, the work performed as part of this dissertation has helped to enhance proteinaceous effector characterization, which, while still early, is thus far the most explored of all PPN infection strategies. This dissertation also explored the possibility of new parasitic mechanisms in the form of secreted exosomes and small RNAs, in an effort to open up new avenues of research. Finally, this dissertation has provided a good example of how high-throughput sequencing has allowed for the exploration of new biological questions relating to PPN signaling.
APPENDIX A. CHAPTER 2 SUPPLEMENTARY INFORMATION

TABLE OF CONTENTS

Table S1. GLAND4 putative transcriptional activation domain ........................................125
Table S2. GLAND4 N-terminal repeat region histone related blast hits ..........................126
Table S3. Primer list ........................................................................................................126
S1 Fig. Pairwise sequence alignment of GLAND4 from H. glycines and H. schachtii .............................................................................................................................128
S2 Fig. GLAND4 expression in transgenic A. thaliana ......................................................129
S3 Fig. Nematode susceptibility in GLAND4 transgenic A. thaliana ..............................129
S4 Fig. AT3G22600 and AT3G22620 expression in transgenic A. thaliana ..................130
Supplemental Methods: Plant transcriptional assays – vector modifications and cloning information ....................................................................................................................130

Table S1. GLAND4 putative transcriptional activation domain. Comparison of the C-terminal of GLAND4 to characterized activation domains from: AvrXa10 (Zhu et al., 1998), OPAQUE-2 (Lohmer et al., 1991), VP16 (Cress and Triezenberg, 1991), and GAL4 (Laughon and Gesteland, 1984). Acidic residues are bolded. Bulky or hydrophobic residues are underlined.

<table>
<thead>
<tr>
<th>Name</th>
<th>Start</th>
<th>Sequence</th>
<th>End</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLAND4</td>
<td>146</td>
<td>EEEEQTDLAQDEGLEDVMDGEEDGMAEEEQPEIN</td>
<td>180</td>
</tr>
<tr>
<td>AvrXa10</td>
<td>1060</td>
<td>TVMWEQDAAPFAGAADDFPAFN EEEELAWLMELLPO</td>
<td>1094</td>
</tr>
<tr>
<td>OPAQUE-2</td>
<td>49</td>
<td>GDGDMDQOHATEWTFERLLEEALTTSTPPP</td>
<td>80</td>
</tr>
<tr>
<td>VP16</td>
<td>422</td>
<td>GDELRLGDGEVDMTPADALDDFDLEMLGDVESPSP</td>
<td>456</td>
</tr>
<tr>
<td>GAL4</td>
<td>850</td>
<td>GITTGMFNTTTMDDVYNLYLFDDEDTPPPKKE</td>
<td>881</td>
</tr>
</tbody>
</table>
**Table S2. GLAND4 N-terminal repeat region histone related blast hits.** A blastp was performed against the non-redundant protein database using the repeat region from GLAND4. Provided in the table are GenBank accession numbers, descriptions and alignment lengths with percent identities to hits with known or predicted histone related functions.

<table>
<thead>
<tr>
<th>Accession</th>
<th>Description</th>
<th>Percent Identity (% positive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OAY64904.1</td>
<td>Histone H1.2 [Ananas comosus]</td>
<td>20 (70)</td>
</tr>
<tr>
<td>AFI74270.1</td>
<td>Inkeeper histone H1 (Musa acuminata AAA Group)</td>
<td>20 (75)</td>
</tr>
<tr>
<td>XP_009413662.1</td>
<td>PREDICTED: histone H1-like [Musa acuminata subsp. malaccensis]</td>
<td>20 (75)</td>
</tr>
<tr>
<td>XP_004828537.1</td>
<td>PREDICTED: histone H1.5 [Octodon degus]</td>
<td>13 (74)</td>
</tr>
<tr>
<td>XP_017493016.1</td>
<td>PREDICTED: late histone H1-like [Rhaepeotes zephyris]</td>
<td>20 (60)</td>
</tr>
<tr>
<td>XP_004289011.1</td>
<td>PREDICTED: histone H1.2 [Fragaria vesca subsp. vesca]</td>
<td>23 (61)</td>
</tr>
<tr>
<td>KX_086693.1</td>
<td>Core histone macro-H2A.1 [Exaiptasia paliida]</td>
<td>20 (65)</td>
</tr>
</tbody>
</table>

---

**S3 Table. Primer list**

**In Situ Hybridization**

2B01_CDS_F 5’-ATAAAAATGCCTGCCGTTCTTCTTCCTGG-3’
2B01_CDS_R 5’-TTAGTTGATCTCCGGCTGTTCTTCCTCA-3’
HsG4_ISH_F 5’-ACTGAACCCTCCAGTGCTCAAGTTGTAG-3’
HsG4_ISH_R 5’-TAGTTGATCTCCGGCTGTTCTTCCTCA-3’

**Subcellular localization**

HsG4_bait_F 5’-gcgc gaattc AAAGCAGTGAAGAAGGACGGCAAAA-3’
HgG4_N1_F 5’-gcgc gaattc AAAGCAGTGAAGAAGGACGGCAAAA-3’
HgG4_N1_R 5’-atat ggatcc AGTTGATCTCCGGCTGTTCTTCCTCA-3’

**EMSA**

G4RE1_F 5’-TGTGTTTTTCACAACCGGCCAA-3’
G4RE1_R 5’-CCAAGTATGTGATTGTTCAGCAG-3’
G4RE2_F 5’-CGACACCTGTTATTTGCTCAAGT-3’
G4RE2_R 5’-GTGACGTTCTAGCGTTTGTG-3’
GRF7_F 5’-GTGTCCGACCGAAAGTAGCCA-3’
GRF7_R 5’-AGGCACAAGATTGAAGAAAGGC-3’

**Quantitative real-time RT-PCR**
Atactin_For 5’-AGTGGTCGTACAAACCGGTATTGT-3’
Atactin_Rev 5’-GAGGATACATGTGGAACGTGACGATAAG-3’
HsActin_F 5’-AAGGCCAACAGGAAAGATGAC-3’
HsActin_R 5’-TTCATCAGGTAGTCTCGAGGTAC-3’

**Generation of Transgenic Arabidopsis**

HsG4_ChIP_FLAG_F 5’-ACAAGGACGATGACGATAAGAGCAGTGAAAGGACGG-3’
HsG4_ChIP_OX_F 5’-gcatGGATCCatgGACTACAAGGACGATGACGATAAGA-3’
HsG4_OX_R 5’-gtatGAGCTCTtaGGTACTGCTCCGGTTGCTGCTCA-3’
At3G22600_OX_BamHI_F 5’-gcggGGATCCATGAAAATGGAAATGAGGTTTAG-3’
At3G22600_OX_SacI_R 5’-ttgGAGCTCTCAGAGATTGCCATGTAGGA-3’
At3G22620_OX_XbaI_F 5’-gtgTCTAAGAATGTCAGTGAGGTC-3’
At3G22620_OX_SacI_R 5’-gttGAGCTCTCAGAGATTGAGGTCTACG-3’
At3G22620_OX_ApaI_MF 5’-ataGGGCCCCAACATCATGACGACGGAGGAA-3’
At3G22620_OX_ApaI_MR 5’-ataGGGCCCCAACCGTGACGAGGTGTTAA-3’

**Plant Transcriptional Assays**

Gal4DB_F_BamHI 5’-GCggaatccATGAAGCTACTGTCTTCTATCGAACAAGC-3’
Gal4_Gene_stopR_Sacl 5’-TATgagctcCTAGTTATGCGGCCTGTCGAG-3’
LexA_CSac_mut 5’-GCCCCCCCCCTCAAGCTCGGAA-3’
LexA_R_BamHI 5’-TATgagctcCTAGTTATGCGGCCTGTCGAG-3’
LexA_R_EcoRI 5’-TACgaattcCAGCCAGTCGCCGTTGCGAA-3’
G4FL_F_EcoRI 5’-GCGgaattcAAAGCAGTGAAAAAGGACGGC-3’
G4FL_R_salI 5’-GATgtcgacGTTGATCTCCGGCTGTTCTTTG-3’
GUS_471_EcoRI_F 5’-cttgaattcGAGACGGACAAAGTGGACGGC-3’
GUS_471_SalI_R 5’-tagtcgacCACGTGATGGTGATGGTGATG-3’
VP16AD_mut1 5’-TCAGCCTGGGGGACGACGACATTCCACTTAGACGG-3’
VP16AD_mut2 5’-CCCGAGCCGATGTCGCCTTTGGG-3’
Cvect_BamHI_F 5’-tatGGATCCatgGATGACGATAAG-3’
Cvect_EcoRI_R 5’-gagGAATTCGATCTAGAATGACTAGACCCCG-3’
6xgal-1
5’AAGAGCTCGGAGTACTGTCTCTCCTCGGAGTACTGTCTCTCCTCGGAGTACTGTCTCTTCCCGGAGTACTGTCTCCTCGGAGTACTGTCTCTCCTCCTCCCGGAGTACTGTCTCCTCGGAGTACTGTCTCTCCTCCTCCGGAGTACTGTCTCCTCGGAGTACTGTCTCTCCTCCTCGGAGTACTGTCTCCTCGGAGTACTGTCTCTCCTCCTCCGGAGTACTGTCTCCTCGGAGTACTGTCTCTCCTCCTCCGGAGTACTGTCTCCTCGGAGTACTGTCTCTCCT-3’
35S-gal -
5’AAGGATCCAGGATGACTGTCTCTCCTCGGAGTACTGTCTCTCCTCGGAGTACTGTCTCTCCTCCTCCCGGAGTACTGTCTCCTCGGAGTACTGTCTCTCCTCCTCCGGAGTACTGTCTCCTCGGAGTACTGTCTCTCCTCCTCCGGAGTACTGTCTCCTCGGAGTACTGTCTCTCCTCCTCCGGAGTACTGTCTCCTCGGAGTACTGTCTCTCCT-3’
Pmod_F 5’-ctaGGATCCtagAAGCTTAAAGAGCTCGGAGTACTGTCTCCT-3’
**S1 Fig. Pairwise sequence alignment of GLAND4 from H. glycines and H. schachtii.** Expasy translate was used to predict the protein sequence of PCR amplified fragments for the H. glycines and H. schachtii GLAND4 orthologs. The sequences were aligned using EMBoss needle with default settings to show 95% identity between the sequences. ( | ) = matching amino acids, (:) = similar conservative amino acids (share the same physicochemical properties). (.) = non-conservative amino acids. The underline denotes the predicted N-terminal signal peptide (SignalP). The positively and negatively charged amino acids are marked with (+) (*) respectively. Amino acids involved in the three overlapping bipartite NLS predictions are marked with (■).
S2 Fig. GLAND4 expression in transgenic A. thaliana. GLAND4<sup>Sp</sup> expression was measured in 3-week old independent homozygous A. thaliana T3 lines (3-10, 5-6 and 6-1) using qPCR. Data were normalized to Actin as an internal control. The means and SE are reflective of two independent experiments each with three biological replicates and four technical replicates. A t-test was used to determine the fold difference as compared to the mean of the lowest expressing line, 3-10 (set at 1.0), marked as a dashed line. The bars represent the averages ±SE *P < 0.01.

S3 Fig. Nematode susceptibility in GLAND4 transgenic A. thaliana. The numbers of female nematodes were counted on independent homozygous A. thaliana T3 lines (3-10, 5-6 and 6-1) constitutively expressing GLAND4<sup>Sp</sup> 4 weeks post inoculation with infective second-stage juveniles. Data were collected from 2 independent experiments each with 20 plants. The bars represent the averages ±SE. A one-way ANOVA was performed to determine no statistical difference between the Col-0 mean and GLAND4<sup>Sp</sup> expressing plants.
Supplementary methods

Plant transcriptional assays – vector modifications and cloning information

*HsGLAND4*<sup>Sp</sup>, GUS and VP16 were first amplified using sequence specific primers, containing EcoRI and SalI sites respectively and inserted into the pGBKT7 vector. Amplification of Gal4BD and Gal4BD+VP16 was performed using Gal4DB_F_BamHI and Gal4_Gene_stopR_Sacl. To create LexA fusions with *HsGLAND4*<sup>Sp</sup> and GUS they were amplified from pGBKT7 without the GAL4BD using a gene specific forward primer containing an EcoRI site and Gal4_Gene_stopR_Sacl. The fragments were digested with EcoRI and ligated to a LexA containing EcoRI fragment. The LexA fusion proteins
were then inserted into pGBK7 and amplified using LexA_F_BamHI and Gal4_Gene_stopR_Sacl (Supplemental Table S3). The Gal4BD and LexA fusion proteins were inserted into the pCXSN vector (Chen et al., 2009). Due to a low level of transformants when using T/A cloning, the pCXSN vector was modified to allow for insertion in between the 35S promoter and polyA using BamHI and Sacl. First the vector was digested with BamHI to remove the ccdB gene. A sacl site was added by PCR using the BamHI ligated vector as a template with Cvect_BamHI_F and Cvect_EcoRI_R (Supplemental Table S3). The Gal4BD and LexA fusion proteins along with BamHI self-ligated pCXSN vector were digested with BamHI and EcoRI to insert the Sacl containing fragment.

The UAS-35Smin fragment was generated by PCR using 6xgal-1 and 35S-gal oligonucleotides. pCXGFP-P was digested with XcmI and the PCR fragment was inserted upstream of GFP. LexAop was added upstream of the UAS-35Smin::GFP vector by modifying the altering the BamHI site directly upstream of GFP to EcoRI using PCR. The LexAop was amplified from pSH18-34 using LexAop_F and LexAop_R and cloned into the Sall and BamHI sites of the modified UAS-35Smin::GFP vector (Supplemental Table S3). All constructs and vector modifications were verified using sanger sequencing.
APPENDIX B. CHAPTER 4 SUPPLEMENTARY INFORMATION

TABLE OF CONTENTS

Table S1 - BLASTn output when using MH119144 as a query sequence against the *H. glycines* genome

Table S2 - BLASTn output when using the first 22 nucleotides of MH119144 as a query sequence against the *H. glycines* genome

Table S3 - BLAST and Sm motif analysis for existing SL sequences

Table S4 - Secondary structural predictions for the putative SL1 RNA genes

Table S5 - SL trans-spliced transcripts are clustered in the *H. glycines* genome

Table S6 – Enrichment of gene ontology function for genes giving rise to trans-spliced transcripts

Table S7 – Repeats associated with trans-spliced, repetitive genes

Table S8 – Effector genes giving rise to SL trans-spliced transcripts

Figure S1 - Sequence alignment of putative HgSL3 RNAs

Figure S2 - Gene Ontology (GO) biological processes specific child terms for SL trans-spliced transcripts
Table S1 - BLASTn output when using MH119114 as a query sequence against the *H. glycines* genome.

| MH119144 | scaffold_282 | 99.75 | 406 | 1 | 0 | 522 | 927 | 38986 | 39391 | 0 | 775 |
| MH119144 | scaffold_282 | 99.75 | 406 | 1 | 0 | 522 | 927 | 42194 | 42599 | 0 | 775 |
| MH119144 | scaffold_282 | 99.55 | 224 | 1 | 0 | 300 | 523 | 38704 | 38927 | 8.53E-118 | 425 |
| MH119144 | scaffold_282 | 99.55 | 224 | 1 | 0 | 300 | 523 | 41912 | 42135 | 8.53E-118 | 425 |
| MH119144 | scaffold_282 | 100 | 184 | 0 | 0 | 120 | 303 | 38432 | 38615 | 2.22E-96 | 354 |
| MH119144 | scaffold_282 | 98.84 | 173 | 2 | 0 | 131 | 303 | 41651 | 41823 | 1.53E-86 | 321 |
| MH119144 | scaffold_282 | 99.05 | 105 | 1 | 0 | 16 | 120 | 38097 | 38201 | 5.41E-49 | 196 |
| MH119144 | scaffold_282 | 80.77 | 78 | 15 | 0 | 38 | 115 | 41180 | 41257 | 5.53E-9 | 64.1 |
| MH119144 | scaffold_189 | 95.57 | 406 | 14 | 1 | 522 | 927 | 126521 | 126922 | 0 | 667 |
| MH119144 | scaffold_189 | 93.81 | 226 | 12 | 1 | 298 | 523 | 126240 | 126463 | 1.74E-93 | 344 |

# BLAST processed 1 queries

Table S2 - BLASTn output when using the first 22 nucleotides of MH119114 as a query sequence against the *H. glycines* genome.

| MH119144 | scaffold_362 | 100 | 22 | 0 | 0 | 1 | 22 | 21923 | 21902 | 7.11E-05 | 43 |
| MH119144 | scaffold_362 | 100 | 22 | 0 | 0 | 1 | 22 | 23092 | 23071 | 7.11E-05 | 43 |
| MH119144 | scaffold_362 | 100 | 22 | 0 | 0 | 1 | 22 | 23484 | 23463 | 7.11E-05 | 43 |
| MH119144 | scaffold_362 | 100 | 22 | 0 | 0 | 1 | 22 | 23875 | 23854 | 7.11E-05 | 43 |
| MH119144 | scaffold_362 | 100 | 21 | 0 | 0 | 2 | 22 | 22700 | 22680 | 2.69E-04 | 41.1 |
| MH119144 | scaffold_362 | 100 | 21 | 0 | 0 | 2 | 22 | 24652 | 24632 | 2.69E-04 | 41.1 |

# BLAST processed 1 queries
Table S3 - BLAST and Sm motif analysis for existing SL sequences. BLASTn output is provided for *C. elegans* and *G. rostochiensis* SLs as query sequences against the *H. glycines* genome. The final column Sm motif (Y/N) indicates yes (Y) or no (N) as to whether each blast result contains a recognized Sm motif (5’-purine-AU<sub>4</sub>·G-purine-3’).

# BLASTN 2.5.0+
# Query: SLs
# Database:
/var/www/scnbase_blastdb/genome738al.polished.mitoFixed.fasta
# Fields: query acc., subject acc., % identity, alignment length, mismatches, gap opens, q. start, q. end, s. start, s. end, evalue, bit score, Sm motif (Y/N)
# 181 hits found

<table>
<thead>
<tr>
<th>SL</th>
<th>Scaffold</th>
<th>Length</th>
<th>Identity</th>
<th>BLAST Length</th>
<th>BLAST E-value</th>
<th>Sm Motif</th>
</tr>
</thead>
<tbody>
<tr>
<td>SL1</td>
<td>scaffold_31</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_31</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_31</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_2</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_2</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_106</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_106</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_83</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_1</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_1</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_1</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_290</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_290</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_19</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_19</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_19</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_19</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_306</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_280</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>-----</td>
<td>------------</td>
<td>-----</td>
<td>----</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold</td>
<td>Start</td>
<td>End</td>
<td>Orientation</td>
<td>Score</td>
<td>Identity</td>
</tr>
<tr>
<td>-------</td>
<td>-----------</td>
<td>-------</td>
<td>-----</td>
<td>-------------</td>
<td>-------</td>
<td>----------</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_89</td>
<td>100</td>
<td>22</td>
<td>0 0 1 22</td>
<td>285705</td>
<td>285726</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_89</td>
<td>100</td>
<td>22</td>
<td>0 0 1 22</td>
<td>298245</td>
<td>298266</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_89</td>
<td>100</td>
<td>22</td>
<td>0 0 1 22</td>
<td>310734</td>
<td>310713</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_103</td>
<td>100</td>
<td>22</td>
<td>0 0 1 22</td>
<td>140281</td>
<td>140302</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_103</td>
<td>100</td>
<td>22</td>
<td>0 0 1 22</td>
<td>141587</td>
<td>141608</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_32</td>
<td>100</td>
<td>22</td>
<td>0 0 1 22</td>
<td>261239</td>
<td>261260</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_10</td>
<td>100</td>
<td>22</td>
<td>0 0 1 22</td>
<td>842563</td>
<td>842584</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_10</td>
<td>100</td>
<td>22</td>
<td>0 0 1 22</td>
<td>852964</td>
<td>852985</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_54</td>
<td>100</td>
<td>22</td>
<td>0 0 1 22</td>
<td>44321</td>
<td>44342</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_6</td>
<td>100</td>
<td>22</td>
<td>0 0 1 22</td>
<td>967409</td>
<td>967430</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_6</td>
<td>100</td>
<td>22</td>
<td>0 0 1 22</td>
<td>971470</td>
<td>971491</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_38</td>
<td>100</td>
<td>22</td>
<td>0 0 1 22</td>
<td>66250</td>
<td>66271</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_38</td>
<td>100</td>
<td>22</td>
<td>0 0 1 22</td>
<td>77360</td>
<td>77381</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_55</td>
<td>100</td>
<td>22</td>
<td>0 0 1 22</td>
<td>416219</td>
<td>416240</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_64</td>
<td>100</td>
<td>22</td>
<td>0 0 1 22</td>
<td>122831</td>
<td>122852</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_16</td>
<td>100</td>
<td>22</td>
<td>0 0 1 22</td>
<td>354047</td>
<td>354068</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_16</td>
<td>100</td>
<td>22</td>
<td>0 0 1 22</td>
<td>375077</td>
<td>375098</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_16</td>
<td>100</td>
<td>22</td>
<td>0 0 1 22</td>
<td>379391</td>
<td>379412</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_16</td>
<td>100</td>
<td>22</td>
<td>0 0 1 22</td>
<td>389453</td>
<td>389432</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_72</td>
<td>100</td>
<td>22</td>
<td>0 0 1 22</td>
<td>313784</td>
<td>313805</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_72</td>
<td>100</td>
<td>22</td>
<td>0 0 1 22</td>
<td>320058</td>
<td>320079</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_72</td>
<td>100</td>
<td>22</td>
<td>0 0 1 22</td>
<td>345396</td>
<td>345375</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_72</td>
<td>100</td>
<td>22</td>
<td>0 0 1 22</td>
<td>351668</td>
<td>351647</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_15</td>
<td>100</td>
<td>22</td>
<td>0 0 1 22</td>
<td>83280</td>
<td>83301</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_15</td>
<td>100</td>
<td>22</td>
<td>0 0 1 22</td>
<td>452674</td>
<td>452695</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_15</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_15</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_135</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_509</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_486</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_486</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_486</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_289</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_289</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_289</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_590</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_518</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_518</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_493</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_434</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_529</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_405</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_390</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_502</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_502</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_330</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_330</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_330</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_330</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_309</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_265</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_236</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_214</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_214</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_156</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_126</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_126</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_109</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>32syntmer</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>24syntmer</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>2syntmer</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_179</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_163</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_163</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_348</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_235</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_275</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_100</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_109</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_132</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_132</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_132</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_132</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_65</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_56</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_27</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_27</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_48</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_91</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_91</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_44</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_7</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_7</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_7</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_61</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_4</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_111</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_87</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_87</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_92</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_88</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_88</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_538</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_507</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_399</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_386</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_294</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_206</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_206</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_699</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_699</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_699</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_699</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_699</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_699</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_699</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_699</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_699</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_621</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_508</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>-------</td>
<td>--------------</td>
<td>-----</td>
<td>-----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_508</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_495</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_341</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_337</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_323</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_323</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_250</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1</td>
<td>scaffold_220</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SL1b</td>
<td>scaffold_48</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Table S4 - Secondary structural predictions for the putative SL1 RNA genes. RNAfold was performed on all SL1 RNA blast hits with a downstream Sm motif. (.) indicates unfolded nucleotides, ( ) are nucleotides involved in hairpin formation.
Table S5 - SL trans-spliced transcripts are clustered in the *H. glycines* genome. The genome was divided into 50kb bins to analyze the number of consecutive SL trans-spliced genes are present in each bin.

<table>
<thead>
<tr>
<th>Scaffold Number</th>
<th>start position</th>
<th>Number of consecutive SL trans-spliced genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>800000</td>
<td>20</td>
</tr>
<tr>
<td>116</td>
<td>100000</td>
<td>19</td>
</tr>
<tr>
<td>127</td>
<td>0</td>
<td>19</td>
</tr>
<tr>
<td>16</td>
<td>700000</td>
<td>19</td>
</tr>
<tr>
<td>141</td>
<td>100000</td>
<td>18</td>
</tr>
<tr>
<td>466</td>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td>5</td>
<td>800000</td>
<td>18</td>
</tr>
<tr>
<td>116</td>
<td>150000</td>
<td>17</td>
</tr>
<tr>
<td>68</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>95</td>
<td>150000</td>
<td>17</td>
</tr>
<tr>
<td>136</td>
<td>200000</td>
<td>16</td>
</tr>
<tr>
<td>192</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>100000</td>
<td>16</td>
</tr>
<tr>
<td>2</td>
<td>150000</td>
<td>16</td>
</tr>
<tr>
<td>21</td>
<td>50000</td>
<td>16</td>
</tr>
<tr>
<td>10</td>
<td>100000</td>
<td>15</td>
</tr>
<tr>
<td>123</td>
<td>50000</td>
<td>15</td>
</tr>
<tr>
<td>127</td>
<td>100000</td>
<td>15</td>
</tr>
<tr>
<td>167</td>
<td>50000</td>
<td>15</td>
</tr>
<tr>
<td>345</td>
<td>0</td>
<td>15</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scaffold Number</th>
<th>start position</th>
<th>Number of consecutive SL trans-spliced genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>200000</td>
<td>15</td>
</tr>
<tr>
<td>447</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>488</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>5</td>
<td>250000</td>
<td>15</td>
</tr>
<tr>
<td>9</td>
<td>500000</td>
<td>15</td>
</tr>
<tr>
<td>10</td>
<td>750000</td>
<td>14</td>
</tr>
<tr>
<td>127</td>
<td>200000</td>
<td>14</td>
</tr>
<tr>
<td>136</td>
<td>150000</td>
<td>14</td>
</tr>
<tr>
<td>136</td>
<td>50000</td>
<td>14</td>
</tr>
<tr>
<td>154</td>
<td>50000</td>
<td>14</td>
</tr>
<tr>
<td>160</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>171</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>19</td>
<td>400000</td>
<td>14</td>
</tr>
<tr>
<td>210</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>229</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>255</td>
<td>50000</td>
<td>14</td>
</tr>
<tr>
<td>51</td>
<td>450000</td>
<td>14</td>
</tr>
<tr>
<td>7</td>
<td>400000</td>
<td>14</td>
</tr>
<tr>
<td>7</td>
<td>800000</td>
<td>14</td>
</tr>
<tr>
<td>8</td>
<td>900000</td>
<td>14</td>
</tr>
</tbody>
</table>
Table S6 – Enrichment of gene ontology function for genes giving rise to trans-spliced transcripts.

<table>
<thead>
<tr>
<th>ID</th>
<th>Pop total</th>
<th>Pop term</th>
<th>Study total</th>
<th>Study term</th>
<th>Pop family</th>
<th>Study family</th>
<th>n parents</th>
<th>is trivial</th>
<th>p</th>
<th>p adjusted</th>
<th>p min</th>
<th>name</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0016301</td>
<td>16030</td>
<td>480</td>
<td>5062</td>
<td>189</td>
<td>568</td>
<td>234</td>
<td>2</td>
<td>FALSE</td>
<td>3.98E-17</td>
<td>8.83E-14</td>
<td>1.40E-256</td>
<td>kinase activity</td>
</tr>
<tr>
<td>GO:0044167</td>
<td>16030</td>
<td>839</td>
<td>5062</td>
<td>379</td>
<td>236</td>
<td>786</td>
<td>2</td>
<td>FALSE</td>
<td>2.96E-13</td>
<td>6.57E-10</td>
<td>6.04E-07</td>
<td>cellular protein metabolic process</td>
</tr>
<tr>
<td>GO:0017675</td>
<td>16030</td>
<td>1085</td>
<td>5062</td>
<td>424</td>
<td>1566</td>
<td>532</td>
<td>1</td>
<td>FALSE</td>
<td>4.52E-11</td>
<td>1.00E-07</td>
<td>5.75E-07</td>
<td>purine nucleotide binding</td>
</tr>
<tr>
<td>GO:0032553</td>
<td>16030</td>
<td>1079</td>
<td>5062</td>
<td>421</td>
<td>1568</td>
<td>532</td>
<td>1</td>
<td>FALSE</td>
<td>8.35E-11</td>
<td>1.05E-07</td>
<td>1.65E-07</td>
<td>ribonucleotide binding</td>
</tr>
<tr>
<td>GO:0036589</td>
<td>16030</td>
<td>1034</td>
<td>5062</td>
<td>399</td>
<td>1602</td>
<td>547</td>
<td>2</td>
<td>FALSE</td>
<td>2.96E-08</td>
<td>6.56E-05</td>
<td>5.75E-05</td>
<td>purine nucleoside triphosphate binding</td>
</tr>
<tr>
<td>GO:0002661</td>
<td>16030</td>
<td>17</td>
<td>5062</td>
<td>10</td>
<td>251</td>
<td>23</td>
<td>1</td>
<td>FALSE</td>
<td>5.03E-08</td>
<td>1.11E-04</td>
<td>5.75E-07</td>
<td>DNA-dependent DNA replication</td>
</tr>
<tr>
<td>GO:0006414</td>
<td>16030</td>
<td>127</td>
<td>5062</td>
<td>64</td>
<td>517</td>
<td>270</td>
<td>2</td>
<td>FALSE</td>
<td>6.38E-08</td>
<td>1.46E-04</td>
<td>1.72E-152</td>
<td>translational elongation</td>
</tr>
<tr>
<td>GO:0043412</td>
<td>16030</td>
<td>810</td>
<td>5062</td>
<td>328</td>
<td>2428</td>
<td>808</td>
<td>1</td>
<td>FALSE</td>
<td>7.31E-06</td>
<td>1.52E-04</td>
<td>6.42E-256</td>
<td>macromolecule modification</td>
</tr>
<tr>
<td>GO:0003723</td>
<td>16030</td>
<td>191</td>
<td>5062</td>
<td>81</td>
<td>1805</td>
<td>470</td>
<td>1</td>
<td>FALSE</td>
<td>1.36E-07</td>
<td>3.03E-04</td>
<td>6.42E-256</td>
<td>RNA binding</td>
</tr>
<tr>
<td>GO:0000575</td>
<td>16030</td>
<td>2016</td>
<td>5062</td>
<td>775</td>
<td>7387</td>
<td>2495</td>
<td>1</td>
<td>FALSE</td>
<td>1.39E-07</td>
<td>3.08E-04</td>
<td>6.04E-07</td>
<td>cellular_component</td>
</tr>
<tr>
<td>GO:0005198</td>
<td>16030</td>
<td>137</td>
<td>5062</td>
<td>100</td>
<td>6525</td>
<td>2163</td>
<td>1</td>
<td>FALSE</td>
<td>1.63E-07</td>
<td>3.62E-04</td>
<td>3.54E-07</td>
<td>structural molecule activity</td>
</tr>
<tr>
<td>GO:0069747</td>
<td>16030</td>
<td>46</td>
<td>5062</td>
<td>19</td>
<td>370</td>
<td>49</td>
<td>1</td>
<td>FALSE</td>
<td>2.29E-07</td>
<td>5.08E-04</td>
<td>7.43E-04</td>
<td>RNA polymerase activity</td>
</tr>
<tr>
<td>GO:0032591</td>
<td>16030</td>
<td>508</td>
<td>5062</td>
<td>242</td>
<td>2015</td>
<td>775</td>
<td>1</td>
<td>FALSE</td>
<td>6.51E-07</td>
<td>0.00144</td>
<td>6.51E-07</td>
<td>macromolecular complex</td>
</tr>
<tr>
<td>GO:0005513</td>
<td>16030</td>
<td>964</td>
<td>5062</td>
<td>381</td>
<td>4267</td>
<td>1417</td>
<td>1</td>
<td>FALSE</td>
<td>1.66E-06</td>
<td>0.00369</td>
<td>1.09E-06</td>
<td>protein binding</td>
</tr>
<tr>
<td>GO:0043604</td>
<td>16030</td>
<td>213</td>
<td>5062</td>
<td>90</td>
<td>1669</td>
<td>310</td>
<td>2</td>
<td>FALSE</td>
<td>2.46E-06</td>
<td>0.00547</td>
<td>4.79E-03</td>
<td>amide biosynthetic process</td>
</tr>
<tr>
<td>GO:0097367</td>
<td>16030</td>
<td>1081</td>
<td>5062</td>
<td>421</td>
<td>4267</td>
<td>1417</td>
<td>1</td>
<td>FALSE</td>
<td>2.54E-06</td>
<td>0.00562</td>
<td>4.79E-03</td>
<td>carbohydrate derivative binding</td>
</tr>
<tr>
<td>GO:0042578</td>
<td>16030</td>
<td>117</td>
<td>5062</td>
<td>39</td>
<td>348</td>
<td>72</td>
<td>1</td>
<td>FALSE</td>
<td>2.60E-06</td>
<td>0.00576</td>
<td>4.79E-03</td>
<td>phosphoric ester hydrolase activity</td>
</tr>
<tr>
<td>GO:0007140</td>
<td>16030</td>
<td>463</td>
<td>5062</td>
<td>201</td>
<td>4821</td>
<td>1642</td>
<td>1</td>
<td>FALSE</td>
<td>6.86E-06</td>
<td>0.00122</td>
<td>6.86E-06</td>
<td>cellular component organization or biogenesis</td>
</tr>
<tr>
<td>GO:0030529</td>
<td>16030</td>
<td>150</td>
<td>5062</td>
<td>86</td>
<td>1165</td>
<td>483</td>
<td>2</td>
<td>FALSE</td>
<td>2.00E-05</td>
<td>0.00435</td>
<td>1.46E-193</td>
<td>intracellular ribonucleoprotein complex</td>
</tr>
</tbody>
</table>

Table S7 – Repeats associated with trans-spliced, repetitive genes.

<table>
<thead>
<tr>
<th>Percent</th>
<th>Repeatmodeller repeat</th>
<th>repeat annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.75756</td>
<td>Motif:rnd-4_family-65</td>
<td>Unknown</td>
</tr>
<tr>
<td>4.27979</td>
<td>Motif:rnd-4_family-1273</td>
<td>LINE/CRI</td>
</tr>
<tr>
<td>3.73444</td>
<td>Motif:rnd-4_family-352</td>
<td>Unknown</td>
</tr>
<tr>
<td>3.10611</td>
<td>Motif:rnd-5_family-2862</td>
<td>Unknown</td>
</tr>
<tr>
<td>2.64375</td>
<td>Motif:rnd-3_family-228</td>
<td>Unknown</td>
</tr>
<tr>
<td>2.62004</td>
<td>Motif:rnd-3_family-42</td>
<td>Unknown</td>
</tr>
<tr>
<td>2.37107</td>
<td>Motif:rnd_4_family-1152</td>
<td>DNA/TcMar-Tc2</td>
</tr>
<tr>
<td>2.03912</td>
<td>Motif:rnd-4_family-299</td>
<td>Unknown</td>
</tr>
<tr>
<td>2.03912</td>
<td>Motif:rnd-4_family-71</td>
<td>DNA/Mule-MuDR</td>
</tr>
</tbody>
</table>
Table S8 – Effector genes giving rise to SL trans-spliced transcripts.

<table>
<thead>
<tr>
<th>Effector</th>
<th>Total number of genes in the genome</th>
<th>Total number of genes that are SL trans-spliced</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLAND5</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>11A06</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>16A01</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>5D08</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>10C02</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4F01</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>17G06</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>33E05</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>10A06</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>19B10</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>21E12</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2D01</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>33A09</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5A08</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8A07</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>HGSB3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>GLAND11</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>GLAND12</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>GLAND4</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

\[\text{scaffold 362 21} \quad \text{scaffold 362 22} \quad \text{scaffold 362 23} \quad \text{scaffold 362 23} \quad \text{scaffold 362 23} \quad \text{scaffold 362 23} \quad \text{scaffold 362 23} \quad \text{scaffold 362 23} \quad \text{cons} \quad \text{cons} \quad \text{cons}\]

Figure S1 - Sequence alignment of putative HgSL3 RNAs. A multiple sequence alignment of putative HgSL3 RNA genes was performed using T-coffee with default settings. (*) = consensus nucleotide position, pink shading = good nucleotide match, yellow and green shading indicates average and bad nucleotide mismatches respectively, (-) = gap.
Figure S2 - Gene Ontology (GO) biological processes specific child terms for SL transspliced transcripts. The number of proteins represented for each GO biological process identified using Blast2GO.
In chapter 2 of this dissertation, the cyst nematode effector GLAND4 was determined to function as a DNA-binding transcriptional repressor of host genes that are involved in plant defense.

**Reporter gene assays to assess the role of GLAND4 on transcription**

When investigating GLAND4, the initial hypothesis was that GLAND4 functioned as a transcriptional activator, rather than a repressor, of host genes. This hypothesis was based on the highly acidic C-terminus of the protein (Fig. S1, Appendix A) and the similarities between the C-terminus of GLAND4 to a number of characterized activation domains (Table S1, appendix A).

A yeast reporter system was used to test the transcriptional activation capabilities of GLAND4 and a series of GLAND4 truncations. The division of GLAND4 truncations was determined based on the distribution of positive and negative charges as well as the location of the neutral linker region, described in chapter 2 of this dissertation (Fig. S1, Appendix A). GLAND4 24-56 is the lysine rich region with
homology to histones, GLAND4 57-109 is the neutral linker region and GLAND4 110-180 is the highly acidic region. Each of the GLAND4 segments was fused to the GAL4 DNA-binding domain (GAL4BD) and transformed into yeast containing the HIS3 reporter gene downstream of the GAL4 binding element, 6xGAL4UAS (Fig. 1A).

Yeast transformants expressing GLAND4 110-180 were capable of growing on selective media (SD/-Trp/-His) (Fig. 1B). All other GAL4BD-effector transformants as

\[ \text{Figure 1. The C-terminal domain of GLAND4 functions as a transcriptional activator in yeast.} \] (A) A schematic representation of the GLAND4 truncation constructs and the control used in the assay, represented using light blue and dark blue boxes, respectively. Each insert was fused in frame with and downstream of the GAL4 DNA-binding domain, represented in orange. A diagram of the histidine reporter gene can be found in the Matchmaker™ Library Construction & Screening Kits User Manuel (Clontech). (B) Yeast cells transformed with a vector containing the GAL4BD are able to grow equally on tryptophan deficient media (SD/-Trp), irrespective of a fusion protein (left panel). Yeast cells containing GLAND4 110-180 are able to grow on media deficient lacking both tryptophan and histidine (SD/-Trp/-His) due to activation of the histidine reporter gene (right panel).
well as the negative controls GAL4BD only and GAL4BD-LaminC were unable to grow on the selective media (Fig. 1B). This indicates that in yeast the C-terminal portion of GLAND4 is capable of transcriptional activation in the absence of the central and N-terminal regions.

A transient expression system was utilized to investigate whether GLAND4 110-180 also functions as an activator in plants. The yeast promoter 6xGAL4UAS and the coding sequence for the UAS binding protein GAL4BD were transferred into separate binary T-vectors and transformed into Agrobacterium tumefaciens for transient expression in Nicotiana benthamiana leaf tissue (Fig. 2A). The 6xGAL4UAS sequence was placed upstream of the green fluorescent protein (GFP) coding sequence, which served as a reporter gene. The levels of GFP expression were monitored during its coinfiltration with individual GAL4BD-effector constructs. The expression levels of hygromycin were used to calculate the amount of infiltrated tissue as it is part of the T-DNA for both the reporter and effector constructs. The reporter was initially coinfiltrated with the GAL4BD alone to establish a basal level of GFP expression when only the GAL4BD is bound to the promoter region (Fig. 2B). To test the sensitivity of the system, the C-terminus of the Meloidogyne incognita effector 7H08 was fused to GAL4BD (GAL4BD-Mi) as this has previously been shown to activate transcription in planta (Zhang et al., 2015). As expected, in the presence of GAL4BD-Mi the level of GFP expression was increased as compared to the basal level (Fig. 2B). An increase in GFP was also observed when GAL4BD was coinfiltrated with GLAND4 110-180, which is consistent with the function of GLAND4 110-180 in yeast (Fig. 2B). Surprisingly the expression of GFP in the presence of full-length GLAND4 resulted in a 2-fold reduction
of expression, as compared to the basal level (Fig. 2B). The finding that full-length GLAND4 reduced GFP expression was what provided the motivation to extensively test if GLAND4 functioned as a transcriptional repressor in chapter 2 of this dissertation.

Figure 2. GLAND4 represses basal gene expression in planta. (A) A schematic representation of the reporter and effector constructs used in the assay. The yeast promoter 6xGAL4UAS was fused upstream of the reporter gene GFP in a binary T-vector. (B) Coinfiltrations of GAL4BD-GLAND4 fusion proteins with the 6xGAL4UAS::GFP reporter. qPCR was used to quantify GFP expression which was normalized to the expression of hygromycin as control for infiltrated tissue. Data presented are the result of two independent experiments each with three biological replicates and four technical replicates. A t-test was used to identify means that were significantly different than the effector construct containing only the GAL4BD, set at 1.0. *P < 0.01 and **P < 0.0001.

A yeast two-hybrid screen to identify GLAND4-host protein interacting candidates

The functional role of numerous cyst nematode effectors has been aided by the use of yeast two-hybrid screens to investigate the potential effector-host protein interactions that occur during infection (Hamamouch et al., 2012; Hewezi et al., 2008;
Hewezi et al., 2010; Hewezi et al., 2015; Lee et al., 2011; Patel et al., 2010; Pogorelko et al., 2016). GLAND4 was therefore used a bait protein to identify potential interacting host proteins using a yeast two-hybrid screen. The yeast two-hybrid screen was conducted according to the methods of previously published effector studies and clontech protocols (Hewezi et al., 2008; Hewezi et al., 2010). As part of the yeast two-hybrid, GLAND4 was used as bait for three different Columbia-0 cDNA libraries that were generated from root tissue harvested at 4, 7 and 14 days post-inoculation. These particular time points serve as representation of the host proteins that are present during each of the three infective juvenile life stages.

The yeast two-hybrid screen using GLAND4 as bait identified eight positive

<table>
<thead>
<tr>
<th>A. thaliana gene ID</th>
<th>Description</th>
<th>Prey clones</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>4 DPI</td>
</tr>
<tr>
<td>AT4G25340</td>
<td>Encodes a member of the FKBP-type immunophilin family that functions as a histone chaperone. Binds to 18S rRNA and represses its expression.</td>
<td>0</td>
</tr>
<tr>
<td>AT1G22920</td>
<td>AIH1 encodes a protein similar to IAB1, a specific mammalian coactivator of AP-1 transcription. Encodes a subunit of the COP9 complex that is involved in protein de-ubiquitination. Plants with mutations in CSN5A and CSN5B have a de-etiolated phenotype.</td>
<td>0</td>
</tr>
<tr>
<td>AT3G04730</td>
<td>early auxin-induced (IAA16)</td>
<td>0</td>
</tr>
<tr>
<td>AT5G07340</td>
<td>Calreticulin family protein; FUNCTIONS IN: unfolded protein binding, calcium ion binding; INVOLVED IN: protein folding</td>
<td>0</td>
</tr>
<tr>
<td>AT3G55520</td>
<td>FKBP-like peptidyl-prolyl cis-trans isomerase family protein; FUNCTIONS IN: FK506 binding, peptidyl-prolyl cis-trans isomerase activity</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 1. GLAND4 yeast two-hybrid interacting candidates. The prey clones column corresponds to the juvenile (J) stage yeast two-hybrid library from which each interacting candidate was discovered.
clones, six of which were identified in the J3 library and two from the J4 library (Table 1). It is noteworthy that the highest number of clones was identified during the J3 stage when GLAND4 expression is highest (Fig. 6C, chapter 2). Sequencing of the positive clones identified five different interacting candidates, one of which, Auxin/INDOLE-3-ACETIC ACID 16 (IAA16) was present four times (Table 1). Interestingly, IAA16 is already known to interact with the cyst nematode effector protein 10A07 (Hewezi et al., 2015).

To test the potential interactions in a more robust manner, yeast co-transformation analyses were performed using both the partial proteins discovered as part of the screen, as well as the full-length proteins, in accordance with Clontech protocols and previously published effector studies (Hewezi et al., 2008; Hewezi et al., 2010). Three of the partial proteins showed a positive interaction only in the presence of GLAND4 (G4) but not in the presence of the negative controls (Fig 3). Testing of the full-length proteins in all three of these cases revealed no growth on selective medium in the presence of GLAND4 (Fig 3). The findings from the co-transformation would suggest that GLAND4 does not interact with host proteins during infection though it should also be considered that GLAND4 may first have to undergo a conformational change, which does not occur using the yeast system, before interacting with a host protein.
Figure 3. Yeast Cotransformation reveals no full length interacting candidates. Yeast cells containing the GLAND4 (G4) bait plasmid and the experimental interacting protein prey plasmid were screened using medium deficient in Leucine and Tryptophan (SD/-Leu/-Trp). Interactions between GLAND4 and the experimental proteins was screened using medium deficient in Leucine, Tryptophan, Adenine and Histidine (SD/-Leu/-Trp/-Ade/-His). Interactions were tested for both the partial proteins identified from the yeast two-hybrid screen as well as their corresponding full-length proteins. Bait plasmid containing no interacting protein (vector) and an unrelated protein (Lamin C) were used as negative controls.
References


Zhang, L., L. J. Davies, and A. A. Elling, 2015, A Meloidogyne incognita effector is imported into the nucleus and exhibits transcriptional activation activity in planta: Molecular Plant Pathology, v. 16, p. 48-60.