ABSTRACT

DIGENNARO, PETER M. Genomic and Structural Analyses of Root-Knot Nematode and Plant Signaling. (Under the direction of Dr. David McK. Bird.)

The study of complex symbioses can be facilitated by considering the mutual constraints imparted by each interacting system. Presented here is the dissection of the Root-Knot Nematode (RKN; *Meloidogyne* spp.) and plant host interactome based on these mutual constraints. RKN is a sedentary obligate parasite of virtually all vascular plant roots, and causes severe crop loss. At the heart of this interaction is the RKN-induced formation of specialized and dedicated nematode feeding sites from host root tissue. To accomplish this degree of plant developmental regulation, RKN must communicate at a level of host biology that is universal to its cosmopolitan range. Concomitantly, the plant signals and pathways utilized by RKN are likely critical to normal plant development. These hypotheses not only allow for a directed dissection of RKN biology, but of plant biology, through the lens of RKN parasitism. Supporting these investigations is a suite of computational tools, as well as completed RKN and host genome sequences. Here, I present evidence consistent with the model that RKN parasitism is based upon communicating with, and responding to, normal host developmental programs at the level of plant signal mimicry, and that host biology is central to our understanding of RKN pathology.

Genomic and Structural Analyses of Root-knot Nematode and Plant Signaling

by
Peter M DiGennaro

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Functional Genomics

Raleigh, North Carolina

2013

APPROVED BY:

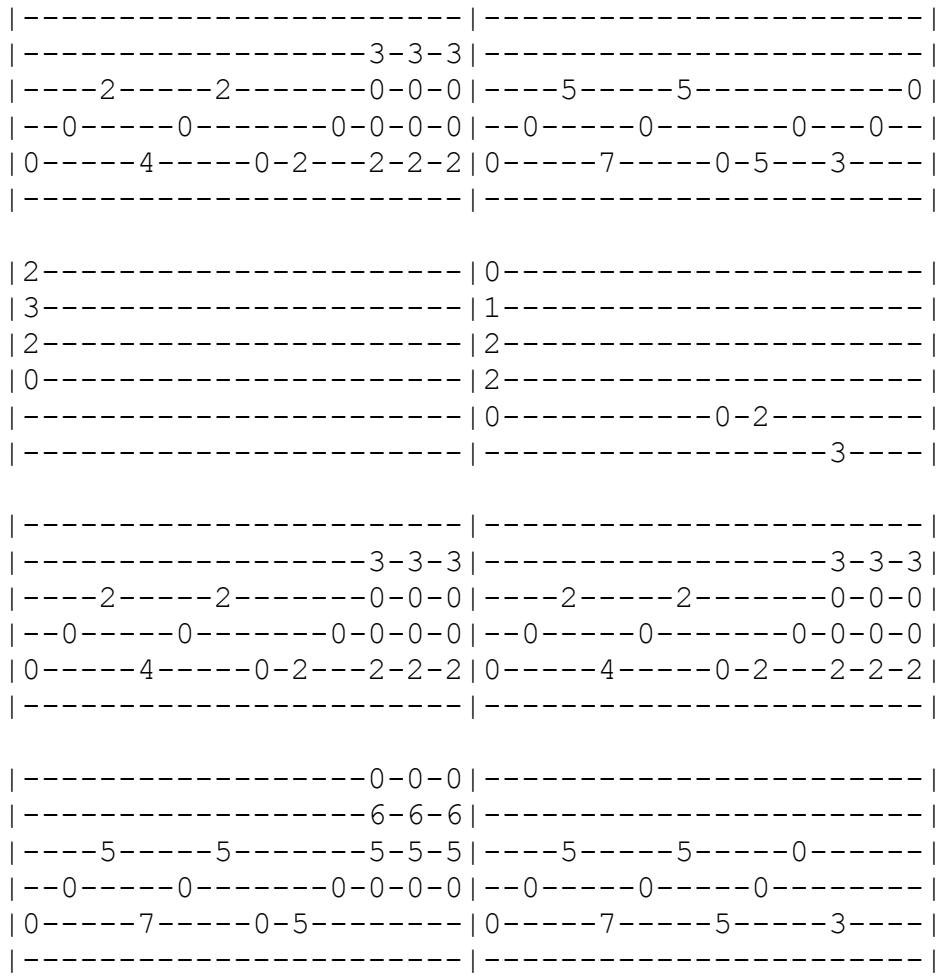
_____	_____
David Mck. Bird Committee Chair	Charles Opperman
_____	_____
Steve Clouse	Laura Mathies

DEDICATION

For Danielle

BIOGRAPHY

I realized my career at North Carolina State University.



ACKNOWLEDGMENTS

For the completion of this dissertation and my degree, I owe a significant debt of gratitude to my major professor Dr. David McK. Bird. His continually innovative vision and dedication to my work instigated the realization of my career. The philosophies of Dr. Bird, and Dr. Charles Opperman went above and beyond material science, reflecting interest in not only my work, but my personal development. I would like to thank Dr. Laura Mathies and Dr. Steve Clouse for serving on my committee--their generous support and perspective supported a comprehensive thesis and broadened the scope of my research. I would like to thank Dr. Benjamin Bobay, Dr. Elizabeth Scholl and Dr. Dahlia Nielsen for making their expertise so readily available, and which made this work feasible. Finally, I would like to thank Stella Chang for her hard work and dedication to the lab and its members.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 2: Structure of a plant peptide hormone and a pathogen-encoded mimic

CHAPTER 3: Structural comparisons of host plant CLE peptides and parasitic root-knot nematode encoded mimics

CHAPTER 4: Invasion of roots by Root-Knot Nematode is Diurnally Biased

CHAPTER 1

Introduction

The pace of plant-parasitic nematology can be described as 'slow.' Antiquated research is still not only relevant, but relatively accurate. This is not for lack of interest, indeed a large body of literature concerning plant parasitic nematodes (PPN) exists; nor is it for lack of scientific merit, as PPN cause billions in agricultural losses worldwide every year (Chitwood, 2003). Rather, the lack of transformational research is the result of the refractory nature of basic PPN biology, specifically for the major obligate sedentary parasite of roots, root-knot nematodes (RKN: *Meloidogyne* spp.).

Begin the genome sequencing era: The completion of the RKN genome (Opperman et al., 2008) permits comprehensive investigations into the genetic armory available to initiate and maintain complex symbioses with host plants. This dataset provided the basis for the research presented here and was able to reconcile previous hypotheses within the field. It is my opinion that, among few others (e.g. the *M. hapla* genetic map (Opperman et al., 2008)), the completion of this data set constitutes the largest advance in accelerating the rate of transformational PPN research. To understand the implications of an available genome and its impact on my research, a brief expose of RKN and host biology and the molecular basis for PPN parasitism is required.

Meloidogyne spp. probably parasitize the roots of all species of higher plants, and possibly lower plants as well such as ferns. In terms of economic importance, RKN infection has been estimated reduce up to 5% of global crop yield pre-harvest (Sasser and Carter, 1985). This impact is at least partly the result of the nematodes cosmopolitan range and high fecundity. Each RKN female may lay up to 1,000 eggs which can persist in the soil and most of which likely infect the same plant. Eggs hatch in the soil as developmentally arrested second stage juveniles (J2, or L2) and have been speculated to be analogous to *Caenorhabditis elegans* dauer larvae. Mechanical and enzymatic penetration of host root, typically near the root apical meristem by J2 begins the endoparasitic stages, where the nematode will complete the rest of its life cycle. Importantly, all nematode life stages post- plant penetration occur in the host apoplast. Once the nematode has migrated to the vascular cylinder, it becomes sedentary, losing its somatic musculature and its vermiform shape. Vascular parenchyma cells near the head of the nematode then undergo developmental reprogramming, finally becoming nematode feeding sites known as Giant Cells (GC). GC are large cells with numerous nuclei, the result of multiple rounds of karyokinesis without cytokinesis. Thickened cell walls and an increased number of organelles are also classical characteristics of GC. In addition, the cortical cells surrounding the inflicted vascular tissue become swollen from hyperplasia. This mass of cell division creates the visible galls, or knots, characteristic of RKN infection. Multiple plant pathways appear to be involved in a successful RKN infection; generally, they are assumed to be i) avoidance and/or inhibition of host defense responses, ii) generation of feeding sites from vascular tissue (GC) and iii) induction of cortical cell hyperplasia.

Also mentioned intermittently throughout the body of this dissertation are the genera *Heterodera* spp. and *Globodera* spp., collectively known as the cyst nematodes (CN). Although there are many superficial similarities between CN and RKN, here the important distinctions will be discussed as they pertain to the dissection of PPN-host interactions. CN are of great agricultural importance, although have a much restricted host range compared to RKN. Like RKN, CN are sedentary obligate plant parasites that induce the formation of feeding sites from host tissue. CN feeding sites form from the coalescence of multiple adjacent pericycle cells, this is distinct from RKN feeding site precursors (vascular parenchyma) and development. In addition, although CN reside in the host apoplast, they migrate *through* plant cells. Electron microscopy and immunochemistry have been used to demonstrate stylet puncturing of syncytial cells (Wyss and Zunke, 1986) and direct secretion of nematode proteins into host cells through the stylet (Elling et al., 2007). Yet, as with RKN, the molecular basis for parasitism remains largely unknown. Once assigned a close phylogenetic relationship with *Meloidogyne* spp. (previously classified as *Heterodera marioni*), more comprehensive phylogenetic analyses indicated an ancient divergence event between CN and RKN (Holterman et al., 2008). Thus, the superficial similarities between these genera (e.g. juvenile morphology, feeding behavior) are likely the result of convergent evolution. Indeed, plant parasitism by nematodes is surmised to have independently arisen at least three times (Baldwin et al., 2004; Meldal et al., 2007). This should caution against assuming the dialog between parasite and host during compatible interactions is similar between these distinct genera. On the other hand, a major theme throughout this work is that

host biology places constraints on parasitic ability, which likely restricts the effective armory of PPN.

Understanding the nature of, and mechanisms by which RKN manipulate plant cell development to form feeding sites has been the focus of concerted efforts for over a century. First postulated by Linford in 1937, nematode secretions have been at the center of these investigations. Since then, a large amount of effort has been concentrated on the nematodes esophageal gland cells and their secreted proteins. The gland cells' presumptive role in feeding site formation is not without basis (Bird and Suarer, 1967), yet many of the putative secreted proteins remain to be experimentally confirmed, and functional classifications remain lacking. Targeting esophageal gland cells for possible signals involved in feeding site formation requires the preconceived notions that these glands are the principal source of signals, and that they are secreted by traditional pathways and contain an amino-terminal secretion signal sequence. To date, the majority of molecules attributed to the host-parasite relationship have been individually cloned by elegant but laborious methods following this hypothesis (Ding et al., 1998; Smant et al., 1998; de Boer et al., 1999; Rosso et al., 1999; Gao et al., 2002; Huang et al., 2003; Qin et al., 2004; Huang et al., 2006a). To characterize a single gene's function, previous approaches have included targeted gene silencing by RNAi (Hewezi et al., 2008, 2010; Hamamouch et al., 2012), localization of gene products *in situ* and *in planta* (most recently and notably, Elling et al., 2007 and Jaoannet et al., 2012), and yeast-two hybrid analyses (Huang et al., 2006a). Together, these methods may provide some insight into a single gene's function, although the entire picture is certainly not represented

due to limitations on each of the technologies. Specifically, RNAi is not robust in all RKN species, and certainly fails in the face of large functionally redundant gene families; *in situ* and *in planta* localization of nematode gene products is often marred by contrived constructs and lack of proper controls for non-specific immunochemistry; and yeast-two hybrid analyses can be fraught with false-positives. As each of these technologies and strategies evolve, they will certainly become more fruitful and robust and will continue to provide novel insight into the molecular basis of RKN parasitism.

No meaningful hypothesis is made without some preconceived notion of the system being studied. Instead of focusing on parasite biology, *viz*. RKN esophageal gland cells and their products, I focused on the current understanding of plant biology and cell development. Assuming RKN parasitism of plants is constrained by the host's ability to recognize nematode products, an understanding of normal plant development may guide the investigations of the signals used by RKN to establish feeding sites.

An abundance of literature exists detailing the interplay of the secondary metabolite plant hormones (abcisic acid, auxin, brassinosteroid, cytokinin, ethylene, gibberellic acid, jasmonic acid, nitric oxide, salicylic acid, and strigolactone), their roles in plant cell development, as well as their perturbation by RKN and CN. The most notable hormones with obvious roles in normal plant development and PPN interactions are auxin, cytokinin and ethylene. Changes in host auxin level are one of the earliest responses to nematode infection. This may be due to localized host defense responses (Phillips et al., 2007), or, more likely, due to changes in the

host's auxin perception and sensitivity. Support for the later comes from the demonstration that expression of the general transcription factor *WRKY23* is rapidly induced upon nematode infection, independent of auxin; *WRKY23* contains four auxin-regulatory elements (Grunewald et al., 2008). This spike in expression may be the result of a nematode signal usurping host gene expression to augment host hormone levels and thus cellular development. Low molecular weight CN compounds have been shown to stimulate proliferation of tobacco protoplasts in the presence of auxin, which is indicative of a heightened sensitivity to this and other plant hormones. Ethylene levels are also responsive to PPN invasion, although once considered a biotic stress response (Glazer et al., 1983; Glazer et al., 1985). Functional analyses of ethylene production and signalling pathways later revealed the hormone to be essential to nematode feeding site formation (Goverse et al., 2000; Wubben et al., 2001). Ethylene has also been implicated in the formation of CN feeding sites, specifically, cell wall modifications permitting solute exchange between the feeding site and the surrounding vascular tissue (Goverse et al., 2000). There is also a fair amount of intrigue around the role of cytokinin in the PPN interaction. Its role in normal plant cell development, particularly cell cycle regulation and division, has certainly aided its continual relevance to investigations into RKN and CN parasitism. Direct application of exogenous cytokinin has failed to show an influence on RKN infection (Dropkin et al., 1969), but does result in the loss of RKN resistance conferred by the *Mi* gene in tomato. The main intrigue around cytokinin lies around the study that demonstrates RKN producing biologically active cytokinin (Bird and Loveys, 1980; De Meutter et al., 2003). The role of this activity in the parasitic interaction remains unclear. During the initial stages of RKN

feeding site development, endogenous cytokinin levels are increased, which likely map to the vascular parenchyma cells destined to become GC (Lohar et al., 2004). Indeed, the resistance conferred by the cytokinin sensitive *Mi* gene is restricted to the initial period of GC induction (Dropkin et al., 1969).

In addition to these classic lipophilic hormones, plant peptide hormones (*pph*) have recently been recognized as significant plant cell developmental regulators. Interrogation of plant genomes reveals large numbers of trans-membrane receptor-like kinases (RLK) and receptor-like proteins (RLP) that represent potential targets for hormone ligands. At least 235 leucine-rich repeat RLK are encoded within the Arabidopsis genome (Sharma et al., 2003), implying numerous opportunities for peptide signaling to regulate the many diverse aspects of plant development attributed to *pph*. However, deducing the role of *pph* in plant tissue and cell development is not straightforward. Over-expression analyses and bioassays often provide contradictory results, which include (but are in no terms limited to) pleiotropic phenotypes (Strabala, 2008). This may be due to ectopic applications and peptide concentrations that lead to neomorphy (new function), hypermorphy (exaggerated function) or antimorphy (opposing function). These results may be due to sensitive and concentration dependent kinetics of the potent ligands (Strabala, 2008).

The RKN-host system provides a valuable to tool to investigate plant biology; primarily because RKN perturb normal plant cell development. It was first proposed by Dropkin and Boone in 1966 and then formalized by Bird in 1996 that the signals employed by RKN to

modulate host pathways may be mimics of host regulatory molecules. In addition, there are likely indirect levels of communication, such as plant circadian and diurnal patterns that influence growth and development, which may be responsive to, and affect nematode biology. Here, I assume that the evolutionary arms race between plant and parasitic nematode has reached a point where nematode signals and the targeted endogenous pathways are intertwined such as to be indistinguishable. To advance the state of plant parasitic nematology, I used available tools to investigate plant developmental regulation through the lens of RKN pathology as it pertains to RKN pathology.

CHAPTER 2

Structure of a plant peptide hormone and a pathogen-encoded mimic

ABSTRACT

CEP regulatory peptides influence root development and architecture in vascular plants. Plant-parasitic root-knot nematodes (RKN; *Meloidogyne* spp.) also manipulate root architecture to establish feeding sites. Genome analysis revealed candidate CEP genes encoded within RKN genomes. Ergo, we hypothesize that CEP play a role in RKN pathogenesis. We used genome-mining, expression analysis and bioassays to establish RKN encode functional CEP-like mimics. NMR-deduced tertiary structures and molecular dynamic simulations were used to investigate the role of CEP in RKN pathology. *M. hapla* variably expresses a genome-compartmentalized 12-member family of CEP-like peptides. Application of synthetic RKN-encoded CEP peptides to host roots suppresses lateral roots and induces galls, which is a hallmark of RKN pathology. NMR structural analysis revealed shared features implicated in general and specific receptor binding; these findings were augmented by molecular dynamic simulations. *M. hapla* encode and express biologically active CEP-mimics that share structural features with host CEP. Based on molecular dynamic simulations, we depict plant CEP acting in multiple tissues and pathways compared with the more rigid and compact nematode CEP, which likely acts in the singular niche of the RKN feeding site. Our analyses implicate CEP as a core signal in the plant-nematode interaction.

INTRODUCTION

Like all organisms, plants interact dynamically with their biotic environment, which includes attack by highly adapted pathogens. Amongst the most sophisticated of such pathogens are plant-parasitic nematodes, which collectively cause as much as $125 billion in annual crop loss worldwide (Koenning et al., 1999; Chitwood, 2003). Much of this damage is caused by sedentary endoparasitic forms, particularly the ~80 species of root-knot nematode (RKN: *Meloidogyne* spp.). RKN hatch in the soil as second-stage larvae and are attracted to roots (Bird, 1959; Riddle and Bird, 1985), which they penetrate near the tip. Inside the root, RKN migrate within the apoplast (Wyss and Grundler, 1992) to the vascular cylinder where the development of a novel plant cell type, termed a Giant Cell (GC), is induced. GC are the essential and, apparently, sole nutritive source for the nematode (Bird and Loveys, 1975). GC formation, coupled with expansion and proliferation of adjacent plant pericycle and cortical cells results in the characteristic root-knot. More than 2,000 plant species are documented as being RKN hosts, although individual isolates may have a more restricted range (Sasser, 1980). All RKN elicit stereotypic GC, implying that the parasite interacts with its host at a level of plant developmental control that is broadly conserved across vascular florae.

It is a long-standing hypothesis that nematode secretions are directly involved in the parasitic interaction (Linford, 1937), and, as recently reviewed (Hewezi and Baum, 2013), the identity of numerous individual proteins in RKN pharyngeal glands and secretions has been established (Ding et al., 1998; Smant et al., 1998; de Boer et al., 1999; Rosso et al., 1999; Gao et al., 2002; Qin et al., 2004; Huang et al., 2006a; Xue et al., 2013). Apart from the well-

studied pharyngeal glands, other *Meloidogyne* cells, including amphidial neurons, have been observed to secrete proteins *in planta,* directed into the host apoplast (Vieira et al., 2011). Intriguingly, included in the RKN and cyst nematode catalogs of secreted proteins are peptides similar to the CLAVATA-like element (CLE) class of plant peptide hormone (Wang et al., 2001; Gao et al., 2001, 2003; Huang et al., 2003). As secreted peptides (12 to 13 amino acids), CLE execute short and long-range communication between plant cells and tissues (Saur et al., 2011; Mortier et al., 2012). Perception of peptide hormones like CLE occurs in the apoplast, and activates a signal transduction cascade in the receiving cell, often invoking different cell developmental programs depending upon the nature of the cell and receptor (Lee and Torii, 2012). Elegant studies have demonstrated that the cyst nematode-encoded CLE is biologically active *in planta*, and functions as a *bona fide* hormone that appears to play a mechanistic role in establishing the syncytial feeding sites induced by cyst nematodes, signifying a true 'mimic' (Wang et al., 2005; Mitchum et al., 2008; Replogle et al., 2013). However, the situation in RKN is less clear. To date, one molecule exhibiting sequence similarity to a plant peptide hormone has been reported, namely, the *M. incognita* 16D10 protein (Huang et al., 2003; Huang et al., 2006a). The 12-amino acid, secreted domain of 16D10 sequence is similar to that of CLAVATA-3. Interestingly, the authors who first described 16D10 later argue that it functions via direct interaction with a plant transcription factor (Huang et al., 2006b).

Although the biochemistry of only a small number of plant peptide hormones and their receptors has been described in detail, computational sweeps of plant genomes have revealed

numerous candidate peptide hormone families (Matsubayashi et al., 2011). Genetic analysis of RKN infection of *Lotus japonicas* (Lohar and Bird, 2003) had implicated a role for receptors of CLE-ligands in the RKN-plant interaction, and the discovery in plant-parasitic nematodes of genes encoding secreted peptides with striking similarity to CLE motivated us to interrogate the *M. hapla* genome for genes with deduced sequence similarity to other plant peptide hormones.

Extensive genomic resources have been developed for *M. hapla*, including the complete 54 Mb diploid genome (Opperman et al., 2008) with annotation supported by more than a billion mapped Illumina RNA-Seq reads (unpublished data). The deduced secreted (Bellafiore et al., 2008) and total (Mbeunkui et al., 2010) proteome has been confirmed in part by mass spectrometry. Collectively, these resources make *M. hapla* a robust candidate for structure–function comparisons. Because of their role in regulating root cell fate, we focused on the plant peptide hormone family 'CEP' (C-terminal Encoded Peptide). Using mass spectrometry, CEP were originally described as a family of *Arabidopsis* secreted peptides (Ohyama et al., 2008). Expression studies revealed distinct spatial and temporal profiles within the CEP family (Ohyama et al., 2008) and endogenous application of synthetic CEP (14-15 amino acids) was found to influence root growth (Ohyama et al., 2008). Transgenic over-expression of *AtCEP1* resulted in shortened roots, the result of a reduction in root meristem cells. Our previous work established that GC have regulatory overlap with primary and lateral root meristems (Koltai et al., 2001; Lohar and Bird, 2003; Lohar et al., 2004;

Weerasinghe et al., 2005). Collectively, these data suggested the hypothesis that CEP play a role in the host-parasite interaction.

A detailed analysis of CEP function in normal plant development and physiological processes is described elsewhere (Imin et al., 2013; also see Ohyama et al., 2008), and points to possible roles for nematode-encoded CEP as mimics of the endogenous hormones. Unfortunately, a number of technical barriers restrict some of the obvious approaches. For instance, it is not yet possible to transform *M. hapla*. Consequently, the strategy of expressing tagged CEP in the nematode as an approach to identify the specific plant receptor is not feasible. Similarly, a precise method of targeted gene knock-out, such as one based on the TALEN system (Li et al., 2011), remains elusive for *M. hapla*. As an alternative, gene knockdown based on RNAi has been successfully used in RKN (reviewed in Lilley et al., 2012). However, when the target gene is a member of an interactive multi-gene family, in which the individual members differ by only a limited number of residues, interpretation of RNAi phenotypes is confounded by three variables, viz., 1) the degree of genetic redundancy, and hence cross-reactivity (such as competition for a receptor), of individual family members, 2) antagonistic/agonistic interactions between the genes/proteins, and 3) the degree to which each gene responds to the RNAi knockdown (i.e., penetrance). Without being able to measure or even estimate these parameters, interpretation of the knock-down experiments is problematic. Similarly, immuno-localization of nematode secreted peptide mimics *in planta* is likely confounded by the lack of unique, nematode-specific epitopes,

such that an antibody would not be able to distinguish nematode from plant encoded peptide ligand.

Prerequisite for specific function of ligands is that the various peptides exhibit a defined tertiary structure. A defined structure is a structure that exhibits stable inter-residue bonds that precludes it as an unfolded and random coil. With these assumptions, differences in the primary structure (i.e., the sequence and post-translational modifications) ought to be manifest as changes in that tertiary structure. Correlating that structure with biological activity thus leads to a powerful predictive tool. Consequently, I have undertaken such a structural approach. I focused the analyses on the model legume *Medicago truncatula* and root-knot nematode *Meloidogyne hapla*. Both have extensive genomic resources and both are of agricultural importance. Importantly, and unlike *Arabidopsis*, *Medicago* A17 is a robust host for *Meloidogyne* species (including *M. hapla*); certain other *Medicago* ecotypes carry resistance to RKN. Based on expression data and peptide sequence, we chose two CEP proteins: CEP1 from *M. truncatula* (MtCEP1) and CEP11 from *M. hapla* (MhCEP11) for structural analysis. Despite spanning just 15 amino acids, we discovered that both peptides exhibit defined structures, signifying specific function. We discuss how CEP mimics and structure pertains to root-knot parasitism.

MATERIALS AND METHODS

Identification of CEP genes: Genome resources for the root-knot nematodes *M. hapla* and *M. incognita*, the free living model nematode *Caenorhabditis elegans*, and the cyst nematodes *Globodera rostochiensis*, and *Heterodera glycines* were obtained from public repositories, including GenBank and WormBase. Genome assemblies for the migratory endoparasites *Pratylenchus coffeae*, and *Radopholus similis* were kindly made available by Dr. C. H. Opperman. A draft assembly *M. chitwoodi* was prepared from sequence in house. Using these genome resources, I extracted the six-frame translations to discover open reading frames between 30 and 150 amino acids, from ATG to stop, using the program getorf from the EMBOSS package (Rice et al., 2000). SignalP 3.0 (Bendtsen et al., 2004) was then used to search for secretion signal sequences in all resulting ORFs, using both neural network and Hidden Markov Model modes (Nielsen et al., 1997; Bendtsen et al., 2004). Any remaining ORFs that contained >5 cysteine residues were removed from further analysis to exclude peptides with multiple intramolecular disulfide bonds, consistent with previous small peptide mining procotols (Ohyama et al., 2008). A custom-made database of ORFs with an identifiable signal sequence was created and searched for the pattern "xfrPTxpGxSPGxGx" developed from the five initial *Arabidopsis* CEP encoding genes reported by Ohyama el al. (2008) using a double-affine Smith-Waterman algorithm (Smith and Waterman, 1981) as implemented in an accelerated TimeLogic hardware environment (TimeLogic DeCypher systems). Resulting matches were hand-curated for conservation of CEP domains as compared to CEP domains found in *A. thaliana*.

RKN CEP gene expression: RKN induced galls were isolated from *Medicago truncatula* cv. Jemalong A17 three weeks after inoculation with *M. hapla* J2 and immediately frozen in liquid nitrogen. Total RNA was isolated from *M. hapla* and *M. truncatula* tissue using the RNeasy Mini Kit (Qiagen) and was processed for sequencing using standard Illumina protocols. Following similar protocols, RNA was also collected from *M. hapla* egg, J2 and whole root tissue of *M. truncatula* containing RKN 1, 2, 4, 5, and 7 days post inoculation with J2. TopHat (Trapnell et al., 2009) and Bowtie (Langmead et al., 2009) were used to align reads to the *M. hapla* genome. *M. hapla* CEP expression was then measured as a percentage of all *M. hapla* CEP reads mapped, or as the percentage of all *M. hapla* reads. Quantitative RT-PCR was used to confirm expression for those *MhCEP* genes for which unequivocal primers could be designed, namely *MhCEP4* and *MhCEP10*.

Peptide synthesis for NMR and bioassays: The active 15 amino acid ligand domains for MhCEP11 (AFRPTAPGHSPGVGH) and MtCEP1 (AFQPTTPGNSPGVGH) were deduced from the appropriate genome sequence, and synthesized to >98% purity with the known modifications of hydroxylation of Pro4 and Pro11, determined by mass-spectrometry (Ohyama et al., 2008) and confirmed for full peptide function (Imin et al., 2013). Syntheses were performed by ChiScientific (Boston, MA, USA), and by the Biomolecular Resource Facility at Australian National University. Synthetic peptide purity was validated by mass-spectroscopy and HPLC. Peptide function was assayed on *M. truncatula*, cv. Jemalong A17 plants using standard methods (Ohyama et al., 2008).

NMR spectroscopy: Structure experiments used a 16.4 T spectrometer at controlled temperature (298.15 ± 0.1 K) on a Bruker Avance 700 MHz equipped with a TCI cryogenic probe. Samples at 4 mg/ml peptide were dissolved in 90% H_2O, 10% D_2O (v/v). DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) was used as an internal reference standard. Two-dimensional natural abundance ^{15}N-heteronuclear single-quantum coherence (HSQC), ^{13}C-HSQC, ^{1}H-^{1}H total correlation spectroscopy (TOCSY) (Bax, 1969) and ^{1}H-^{1}H nuclear Overhauser effect spectroscopy (NOESY) (Wüthrich, 1986) spectra were recorded with standard pulse sequences. The mixing times for ^{1}H-^{1}H TOCSY were 30 and 80 ms and the ^{1}H-^{1}H NOESY spectra were collected with mixing times of 50 and 300 ms with a 3.0 s recycling delay. In all experiments, the water peak was suppressed by presaturation. All the NMR spectra were processed with NMRPipe (Delaglio et al., 1995) and analyzed with the NMRViewJ program (Johnson, 2004).

Peptide structure determination: NMRViewJ software was used to integrate NOE cross-peaks (Johnson, 2004). Using default parameters, ARIA (Ambiguous Restraints for Iterative Assignment - version 2.3) was used to perform structure calculations (Habeck et al., 2004; Linge et al., 2003). The 20 lowest energy structures from each iteration were used for NOE assignment and used for each subsequent iteration. After eight iterations, the 20 lowest energy structures were refined by a molecular dynamics step in explicit solvent (water). From this, the 10 structures with the lowest energy were considered as accurately characterizing the peptide structures and are presented. Representation and quantitative analysis of the calculated structures were performed using PyMOL (Version 1.5.0.4 Schrödinger, LLC).

Molecular dynamics simulation: Molecular Dynamic (MD) simulations were performed with the GROMACS 4.4.5 software package using the AMBER 99sb-ildn force field and the flexible SPC water model. The initial structure was immersed in a periodic water box of cube shape (1 nm thickness) and neutralized with 3 and 1 Cl- counterions for MhCEP11 and MtCEP1, respectively. Electrostatic energy was calculated using the particle mesh Ewald method. Cutoff distances for the calculation of the Coulomb and van der Waals interaction were 1.0 and 1.4 nm, respectively. After energy minimization using a steepest decent method, the system was subject to equilibration at 300 K and normal pressure for 100 psec under the conditions of position restraints for heavy atoms and LINCS constraints for all bonds. The system was coupled to the external bath by the Parrinello-Rahman pressure and temperature coupling. The final MD calculations were performed under the same conditions except that the position restraints were removed and the simulation was run for a full nano second solely to characterize the peptides' structures. The results were analyzed using the standard software provided by the GROMACS package; g_hbond and g_gyrate.

RESULTS

Root-knot nematodes encode CEP-like proteins: Mining public genomic and in-house datasets revealed that the CEP 15 amino acid motif (Fig. 1) is restricted to vascular plants, with one notable exception; the obligate plant-parasitic root-knot nematodes encode a family of proteins with striking sequence similarity to plant CEP (Fig. 1). In no instance, including soybean- and potato-cyst nematode (*Heterodera glycines* and *Globodera rostochiensis*), the migratory plant parasitic nematodes (*Radopholus similis* and *Pratylenchus coffeae*) or the

free living *C. elegans*, were *CEP* genes detected. I numbered the 12 *M. hapla* CEP in the order they appear in the genome, rather than based on the extent of the sequence similarity to plant peptide hormones. In *M. hapla*, *MhCEP1-9* are tightly linked together at one locus on contig 363, while *MhCEP10-12* were clustered together nearby on contig 368; all *MhCEP* genes are located within 140kb on chromosome 9 (www.hapla.org). Both *CEP* clusters occur in an otherwise gene-poor region of an otherwise compact genome.

Comparison of the deduced CEP-like proteins encoded by RKN and Medicago revealed, in addition to the CEP domain, a well-defined secretion signal sequence. In nematodes the signal sequence is immediately adjacent (MhCEP3, MhCEP6 and MhCEP7) or within a few amino acids of the putative ligand domain. This is in contrast to plant CEPs which encode an obvious pro domain between the signal sequence and the ligand domain. As is the case for plant CEP, some RKN-encoded CEP have additional amino acids carboxyl to the 15mer active ligand (Fig. S1).

CEP genes are expressed in RKN, and the peptides they encode are biologically active: In the absence of prior spatio-temporal data on the expression of *MhCEP* genes, we designed two complimentary experiments to broadly sample RKN transcriptional space before and after penetration of host plants. First we undertook a highly replicated RNA-Seq study of mRNA isolated from the knots present on *M. truncatula* roots 3 weeks after inoculation with *M. hapla*. Under our conditions, this time-point captures the nematode actively feeding from established GC, but before the onset of egg-laying. Independently sequencing the 115

biological replicates yielded more than 4.4 billion reads that could be mapped onto either the *Medicago* or *Meloidogyne* genomes. Within this data set, transcripts from all 12 *MhCEP* were detected, although at 3 weeks post inoculation, expression of *MhCEP4, 6,* and *9* accounted for 18.6, 43.2 and 19.3% respectively, of the total CEP reads. To minimize false positives, we scored only sequence reads which uniquely mapped to a CEP coding region. Each RKN *CEP* gene consists of a single exon with an average length of 174bp.

To obtain a more global view of RKN CEP expression, and to confirm our findings from the 3-week analysis, we examined the expression of *MhCEP* genes during multiple nematode life stages. Independent sequencing runs yielding 6.19E+07 *M. hapla* RNA-Seq vreads, were conducted on RNA from *M. hapla* eggs, pre-penetration J2, and post-penetration stages sampled at time points up to 7-days post-inoculation. Normalizing the mapped CEP reads to the total number of *M. hapla* reads for each tissue type gave an approximation of the abundance of each CEP transcript for that time point. Reflecting that, whole infected root tissue was collected for RNA sequencing, only 4.5% of all reads were unambiguously mapped and attributed to *M. hapla* gene expression, with the remainder mapping to the host plant. Despite this dilution, the results are consistent with what we had observed at 3 weeks. Expression of *MhCEP9* at 3 weeks, was confirmed at earlier time points as well. Expressed as a percentage of total *M. hapla* reads mapped, *MhCEP9* was expressed in egg (2.20E-05%), J2 (4.09E-05%) and 2 (1.68E-03%), 4 (2.24E-04%), 5 (1.33E-02%), and 7 days post-inoculation (1.87E-03%). Most other *MhCEP* genes are not expressed or more likely, expressed below the level of detection of this biological sample. Quantitative RT-PCR was

conducted on *M. hapla* CEP encoding genes for which unequivocal, specific primers could be designed (*MhCEP4*, *MhCEP10*), these results verified CEP expression during the parasitic interaction (Fig. S2).

We chose *MhCEP11* for further bio-activity and structural analysis based on several criteria. First, it exhibits highest sequence similarity with *Medicago truncatula MtCEP1*. MtCEP1 is the presumptive ortholog of the canonical CEP from Arabidopsis (*AtCEP1*). Second, because of its status there is a disproportionate abundance of functional and expression data on *MtCEP1*. Third, although the steady state transcript levels of *MhCEP11* are relatively low at 3 weeks (0.13% of all *MhCEP* mapped reads), *MhCEP11* is expressed by the nematode prior to host invasion, consistent with a role in the host-parasite interaction. Specifically, expression in eggs and J2 was 1.32E-05% and 4.5E-05% *M. hapla* reads mapped, respectively.

High purity peptides (>95%) were synthesized for use in bioassays to investigate the role RKN-encoded CEP ligands might play in the host-parasite interaction. Strikingly, *Medicago truncatula* seedling roots treated with the nematode peptide (MhCEP11) exhibited a reduction in lateral root number and gall-like root swellings (Fig. 1); this confirmed RKN CEP functional activity and validated RKN CEPs as mimics of the endogenous ligands.

NMR spectroscopy of nematode and host encoded CEP ligands: The NMR chemical shift assignments were made through a use of the natural abundance ^{15}NHSQC, ^{13}CHSQC, ^{1}H-^{1}HTOCSY, and ^{1}H-^{1}HNOESY experiments. Near complete proton backbone resonances for HN (100%), Hα (93%) and Hβ (100%) were assigned including 100% of NMR assignable side-chain for MhCEP11 (Table 1). For MtCEP1, near complete backbone resonances for HN (100%), Hα (100%), and Hβ (100%) were assigned including 100% of NMR assignable side-chain for MtCEP1 (Table 2). The peptide's structural quality and validation are shown in Table 3. The solved structures were obtained from a total of 50 or greater NOEs, have no NOE violations greater than 0.3 Å and display root mean square deviations of ~1.5 Å over all backbone atoms. Each peptide was characterized by nearly the same amount of inter-residue NOEs as intra-residue NOEs, ~25-30. Furthermore, both peptides have Ramachandran space for each residue >90% in generously allowed space or better.

The peptides display amide (HN) chemical shifts centered around 8.4 ppm (Fig. 2). In the presence of 10% D_2O, both peptides retained HN resonances, indicating the HN atoms are involved in interactions that preclude them from exchanging with the D_2O in the NMR solvent (Fig. 2). When comparing the TOCSY and NOESY spectrum for each peptide there are observable inter-residue interactions taking place (Fig. 3). Fig. 2 displays an overlay of the 80ms TOCSY (black peaks) and 300ms NOESY (red peaks) spectrum for each peptide. There are inter-residues interactions observed throughout the full spectrum (0-12 ppm; red peaks do not exactly overlay black peaks) which lead to the >50 NOEs assigned for the structures.

The NOEs used to characterize the structures are summarized in Fig. 4a,b. The NOEs are observed throughout the sequence of both peptides, not just within the core of the peptide. Furthermore, there are NOEs observed to the N- and C-terminus. For MhCEP11 there are a total of 11 NOEs in the termini (residues 1-2 and 14-15) 5 of which are inter-residue, while for MtCEP1 there are a total of 8 NOEs in the termini (residues 1-2 and 14-15) 3 of which are inter-residue.

The MhCEP11 peptide ensemble has a clearly defined alpha-helical character between residues 9-12 (Fig. 4). In comparison, MtCEP1 contains several beta-turns (residues 1-4, 6-9, and 10-13), and a perceived propensity of resides 8-13 to form alpha-helix (Fig. 4 and Fig. S3). The solution structures of both peptides reveal residue side-chains that are quite dynamic (Fig. 4, Table 3). The flexibility in the side chains is most evident in the amino- and carboxyl-terminal residues of each peptide. When these residues are removed from the root mean square difference (r.m.s.d) calculations, the r.m.s.d is reduced from 1.7 Å to 1.1 Å for MhCEP11 and from 1.2 Å to 1.0 Å for MtCEP1.

Molecular dynamics simulations of solved CEP structures: Molecular dynamics simulations (MD) were run to further characterize the peptides structure and determine hydrogen bonds and compactness. The 1 nanosecond MD simulations were run to solely characterize the solved peptide structures and define hydrogen bonding patterns and compactness which are used to implicate relative functional plasticity or rigidity. Longer simulations are required to be able to describe the peptides dynamics, which is beyond the scope of this study.

Analyzing the hydrogen-bonding pattern over the simulation revealed 11 and 2 hydrogen bonds existing for MhCEP11 (for hydroxylated Pro11 and hydroxylated Pro4, respectively) and 6 and 8 hydrogen bonds existing for MtCEP1. There were a total of 37 hydrogen bonds over the life of the simulation for MhCEP11 and 48 for MtCEP1 (Fig. 5a,b). A pattern for the peptides' structures is evident when looking at the most populated hydrogen bonds. For both peptides, there are hydrogen bonds throughout the sequence, owning to the observed HN resonances as these atoms are involved in interaction that preclude the exchange of D_2O (Fig. 2). Furthermore, MhECP11 and MtCEP1 have hydrogen bonds from amides to side chain groups within each sequence. The radius of gyration (i.e., compactness of the structures) of the peptides was also determined (Fig. 5c). Both peptides have a level of compactness mitigated by a series of hydrogen bonds, salt bridges and hydrophobic interactions. MhCEP11 has an average radius of gyration of 0.55 Å. MtCEP1's compactness is slightly less than MhCEP11 at an average of 0.65 Å.

DISCUSSION

Here we demonstrate that the plant parasitic root-knot nematode uniquely encodes and expresses bioactive mimics of a plant peptide hormone family during the parasitic interaction with a host plant. Further, based on our structural comparisons, we have revealed novel insights into the functional equivalence of endogenous hormones and RKN encoded mimics. In a system largely recalcitrant to genetic manipulation, we show that structural determinations and comparisons offer a supplement to genomic interrogations, particularly for non-model organisms.

CEP are unique to RKN and vascular plants: CEP were first described as a family of bioactive peptide hormones in Arabidopsis (Ohyama et al., 2008). Our mining of the now much larger genomic datasets confirms that observation, with the one notable exception of CEP-like genes encoded within the RKN genomes (Fig. 1; Fig. S1). Although the complete genome sequence of the diploid RKN species, *M. hapla*, predicts approximately 14,500 genes (Opperman et al., 2008), the CEP-like loci had not been predicted by the gene-finding algorithms, likely a consequence of their compact nature and small (15 amino acid) query. Because of the absence of prior annotation of CEP in the RKN genome, we carefully examined the available genome resources for other obligate plant parasitic nematodes, using the very conservative Smith-Waterman algorithm. In no other available nematode genome were CEP-like genes detected. Apart from RKN, extensive bioinformatic analysis detected credible CEP genes in vascular plants only. The absence of credible CEP genes in the cyst nematodes is especially striking, because like RKN, they are sedentary, root endoparasites, and share other parasitism functions with RKN.

RKN bioactive CEP mimics are expressed during the nematode-host interaction: In *Arabidopsis* CEPs are translated as pre-pro-proteins. Cleavage of the pro domain in the apoplast is necessary for activity, presumably serving as an additional regulatory check on the activity of these potent peptides (Ni et al., 2011). By contrast, RKN CEP-like proteins lack an obvious pro domain and are predicted to be secreted directly into the host apoplast as fully active ligands, circumventing host regulatory machinery. Due to limited available epitopes and high sequence similarity to endogenous peptide hormones, direct *in planta*

detection of RKN CEP is currently not possible. Some RKN CEP-like proteins encode additional amino acids carboxyl to the 15mer active ligand (Fig. S1). These residues either have little influence on function, or more likely, are proteolytically removed by the host (Guo et al., 2011). Multiple independent RNA-Seq analyses of *M. hapla* at various life stages confirmed transcriptional activity of *MhCEP*. Different family members probably exhibit different steady state levels of transcript, and multiple expression profiles can be envisioned for members of the RKN CEP gene family including host specificity and dependence upon environmental conditions (Imin et al., 2013).

Given the similarity of RKN CEP-like ligands to plant CEP, I hypothesized that the nematode peptides may play a role in the developmental transitions that occur during infection of plants. The original report on CEP function in plants demonstrated arrested root growth attributed to a reduction in root meristematic cell number (Ohyama et al., 2008). Additional reports demonstrated gall-like root swellings from over-expression of *Medicago truncatula CEP1* (MtCEP1) and exogenous application of a synthetic peptide (Imin et al., 2013). Strikingly, roots treated with the nematode peptide (MhCEP11) exhibited both a reduction in lateral root number and gall-like root swellings (Fig. 1). Due to the functional equivalence of RKN encoded CEP and endogenous CEP ligands, we describe RKN encoded CEP as mimics. It is important to note that a characteristic of parasitism by RKN is hyperplasia of cortical cells surrounding the GC formation of a gall. Intriguingly, *MhCEP9* expression is highest at five days post inoculation, concurrent with the emergence of visible

RKN-induced root galls. The root-assays recapitulate those obtained in parallel experiments on *Medicago* exposed to MtCEP1 and detailed elsewhere (Imin et al., 2013).

Host and RKN CEP 15mer mimics have defined structures: To investigate and compare structure function relationships, we determined the well-defined tertiary structures of MtCEP1 and MhCEP11 using NMR. Superficially, the structures of MtCEP1 and MhCEP11 peptides appear to be quite different. However, they do share several similar features, such as beta turns at the amino- and carboxyl-termini as well as a structured core. Specifically both peptides occupy similar volumes, indicative of well-defined structures. These peptides also have similar arrangements of hydrophobic residues/patches, partial positive, negative and charged regions (Fig. 6). Collectively, these conserved features may point to general receptor recognition properties for this family of peptides.

The MhCEP11 peptide ensemble has the unique structural feature of a clearly defined alpha-helical character between residues 9-12. In comparison, MtCEP1 contains several beta-turns (residues 1-4, 6-9, and 10-13), and a perceived propensity of resides 8-13 to form alpha-helix (Fig. 4); the presence of which would likely limit the peptide's ability to sample conformational space. Like other small peptides, the MtCEP1 and MhCEP11 solution structures reveal quite dynamic residue side-chains. The flexibility in the side chains is most evident in the amino- and carboxyl-terminal residues. When these residues are removed from the r.m.s.d (root mean square difference) calculations, the r.m.s.d is reduced from 1.7Å to 1.1Å for MhCEP11 and from 1.2Å to 1.0Å for MtCEP1. Thus, while the cores of each

peptide exhibit a stable backbone conformation, the ability to sample conformational space is retained by the flexible side chains. The general recognition properties garnered through similar backbone structural motifs allow the side chains to position themselves in arrangements that allow for dynamic movement and could lend themselves to specific functions.

Post-translational modifications have implications for structure and function: Results from *in vitro* bioassays implicate hydroxylation of the proline residues 4 and 11 (on the γ position) as an absolute requirement for full functionality for these peptides (Imin et al., 2013). From the structures we obtained, it is clear these residues are exposed to the solvent which points to a role in receptor binding. However, the conformation of the hydroxyl atoms is distinct for each peptide. While the hydroxyprolines are structurally adjacent in both peptides, they are pointing in opposite directions in MhCEP11, and towards one another in MtCEP1. One possible conclusion to be drawn from the distinct arrangements of these residues is that hydroxylation of either Pro4 or Pro11 is not strictly necessary for function of the peptides. In fact, mass spectrometry of intrinsic AtCEP1 revealed that hydroxylation of Pro4 *in planta* is variable, with non-hydroxylated residues constituting a significant amount of detected peptide (Ohyama et al., 2008). This precludes the hydroxylation of Pro4 as being a prerequisite for correct proteolytic processing of the pro-domain in endogenous peptides (Ohyama et al., 2008). Interestingly, the lack of a pro-domain in CEP mimics encoded by RKN might suggest functional plasticity, indicating that RKN make more ‘general’ ligands,

presumably reflective of the nematodes extensive host range. Additionally, mutagenesis of select amino acids showed that removal of the hydroxyl group on Pro4 of MtCEP1 impairs, but does not eliminate the gall-inducing function of the peptide (Imin et al., 2013).

In the initial description of CEP in *Arabidopsis*, the expression patterns of these genes suggested that the peptides function in diverse tissues (root versus shoot). This is also true for *MtCEP1*. Using the publically available *Medicago truncatula* Gene Expression Atlas (mtgea.noble.org) I examined *MtCEP1* (accession Mtr.7265.1.S1_at) expression in various tissue types (Fig. S4). *MtCEP1* is expressed both in roots and above ground tissue, suggestive of a role in multiple plant developmental pathways. The variable hydroxylation of *Arabidopsis* CEP Pro4 might serve to enhance such functional promiscuity. Alternatively, it may function as a 'dial' of peptide potency during ligand-receptor binding. Aside from mediating receptor recognition, post-translational modifications on peptide hormones are also speculated to function in regulating transport, resistance to enzyme degradation and storage (Matsubayashi et al., 2011).

Molecular dynamics simulations and roles of peptide hormones in plants: To understand the implications of the defined CEP structures and how they might act in host apoplast, we simulated the molecular dynamics (MD) of both peptides in explicit water. Analyzing the hydrogen bonding pattern revealed 11 and 2 hydrogen bonds existing for MhCEP11 (for Pro11 and Pro4, respectively) and 6 and 8 hydrogen bonds existing for MtCEP1. More potential hydrogen bonding events implies the peptide is flexible and able to interchange

hydrogen bonding pairs when in solution. The limited number of hydrogen bonds formed with hydroxylated Pro4 in MhCEP11 is congruent with the hypothesis that it is not required for full CEP function and structure. There were total of 37 hydrogen bonds over the life of the simulation for MhCEP11 and 48 for MtCEP1. Taken together, MtCEP1 was observed to have more hydrogen bonds in total, and of the hydrogen bonds involving hydroxylated proline residues, they were more evenly split compared to those in MhCEP11. When hydroxylated proline residues are not directly interacting, as is the case for MhCEP11, hydrogen bonding between other residues is required to maintain a defined structure. However, the extensive hydrogen bonding network within MtCEP1 implies a more compact functional structure compared to MhCEP11. The radius of gyration, which reflects the compactness of the structures, was also monitored. Both peptides appear to maintain a level of compactness throughout the simulation, suggestive of a folded peptide maintaining its structure through a series of hydrogen bonds and salt bridges. Yet, the two peptides have differences in their overall compactness. The smaller radius of gyration for MhCEP11 is indicative of a more compact structure; this is confirmed through visual inspection of the peptide structures. MtCEP1 is less compact in comparison, which may be another indication of functional plasticity. Considering MtCEP1 is expressed within multiple plant tissues (Fig. S4), it can be surmised that MtCEP1 has the capacity to act in multiple pathways; structural flexibility would facilitates this promiscuity. As a precedent, the canonical CLE ligand, CLV3 has multiple known receptor complexes, and ultimately the function of CLV3 is dictated by the nature and location of the receptor (Gao et al., 2012). MhCEP11 action however, is most likely restricted to the niche of *M. hapla*, *viz.,* the root apoplast near

pericycle and cortical cells. The more rigid structures could possibly reflect an adaptation that targets receptors specific to the limited niche of nematode parasitism.

Table 1. ^{1}H NMR Chemical Shifts of MhCEP11 (Relative to DSS, ppm)

Residue	*HN*	*Ha*	*Hb*	*other*
Ala1	8.082	4.014	1.473	
Phe2	8.649	4.614	3.041	
Arg3	8.212	4.571	1.738	Hg 1.574,1.630 Hd 3.156
Hyp4		4.589	2.357	Hg 2.016 Hd 3.756
Thr5	8.394	4.233	4.133	Hg 1.210
Ala6	8.356	4.603	1.323	
Pro7		4.401	2.283	Hg 1.902,2.027 Hd 3.654,3.797
Gly8	8.528	3.956		
His9	7.961	4.518	3.263,3.128	
Ser10	8.502	4.747	3.838,3.799	
Hyp11		4.61	2.351,2.342	Hg 2.074 Hd 3.867,3.861
Gly12	8.543	3.897		
Val13	8.009	4.129	2.074	Hg 0.893
Gly14	8.504			
His15	8.327	4.739	3.183,3.252	

Table 2. ^{1}H NMR Chemical Shifts of MtCEP1 (Relative to DSS, ppm)

Residue	*HN*	*Ha*	*Hb*	*other*
Ala1	8.024	4.471	1.145	
Phe2	8.627	4.605	3.073,2.346	
				Hg 1.986,1.852
Gln3	8.251	4.602	2.308	He 7.472,6.857
				Hg 2.016
Hyp4		4.6	2.37	Hd 3.791,3.643
Thr5	8.445	4.345	4.186	Hg 1.207
Thr6	8.237	4.607	4.128	Hg 1.214
				Hg 2.073,2.017
Pro7		4.401	2.275,1.905	Hd 3.858, 3.753
Gly8	8.536	3.951		
Asn9	8.265	4.762	2.787,2.720	Hd 7.539, 6.891
Ser10	8.316	4.73	3.826,3.779	
				Hg 2.068
Hyp11		4.589	2.313	Hd 3.851
Gly12	8.437	3.934		
Val13	7.984	4.124	2.081	Hg 0.889
Gly14	8.509	3.893		
His15	8.005	4.513	3.251,3.094	

Table 3. Details and statistics for quality assessment of peptide structures.

	MhCEP11	**MtCEP1**
Total NOES	50	58
Intra-residue	24	26
Inter-residue	26	32
Unambiguous	43	43
Ambiguous	7	15
Average violations per structure		
> 0.5Å	0	0
> 0.3Å	0	0
> 0.1Å	1.4	2.5
Energy (kcal mol-1)		
van der Waals	-88.73 (± 5.48)	-86.51 (± 4.54)
electrostatic	-409.61 (± 44.75)	-515.14 (± 16.55)
Deviations from idealized geometry		
Bond lengths (Å)	0.0036 (± 0.0001)	0.0041 (± 0.0003)
Bond angles (º)	0.564 (± 0.036)	0.630 (± 0.042)
Impropers (º)	1.47 (± 0.23)	1.36 (± 0.52)
Ramachandran Space		
Core	61.1%	38.6%
Additionally	34.4%	50.0%
Generously	3.3%	9.1%
Disallowed	1.1%	2.3%
RMSD (Å)		
Backbone	1.751 (± 0.724)	1.255 (± 0.359)
Heavy atoms	2.437 (± 0.621)	1.959 (± 0.637)

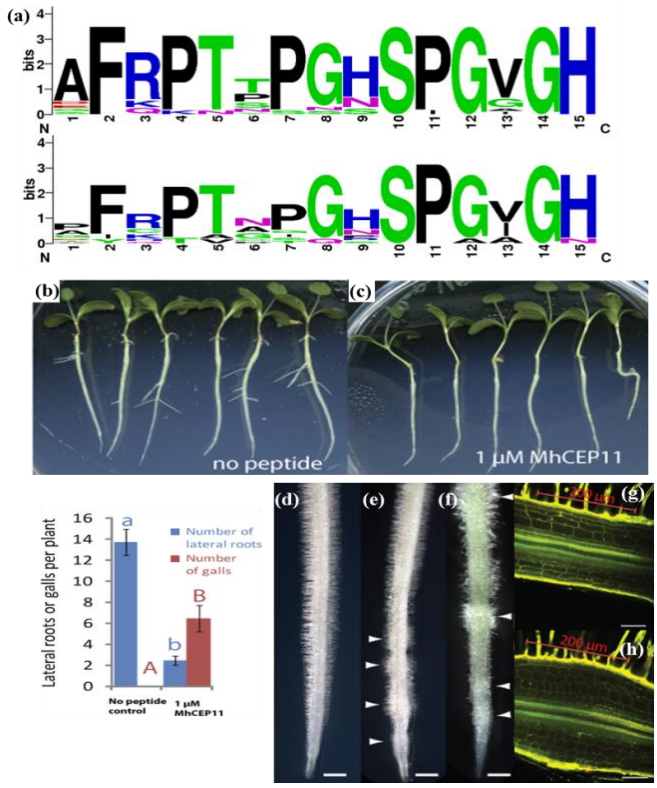


Figure 1. RKN encode and expression bioactive CEP mimics

(a)Logo plots of the active 15mer ligand domains of the 12 *M. hapla* (top) and the 11 *M. truncatula* CEP ligands (bottom). General chemical properties of amino acids are color coded. (b, c) *M. truncatula* roots grown on vertical agar plates in the absence of CEP (b), lateral roots are apparent while the addition of 1µM MhCEP11 (c) suppresses lateral roots and induces gall formation (graph; mean values and standard deviations were calculated from six replicate samples). (d - f) Exogenous application of synthetic MtCEP1 (e) or MhCEP11 (f) to growing *M. truncatula* roots results in the periodic formation of galls and reduction in lateral roots compared to no peptide control (d); scale bar size is 200µM. (g, h) Confocal imaging of control root (g) and MhCEP11 induced root gall demonstrating gall is from hyperplasia of cortical cells (f).

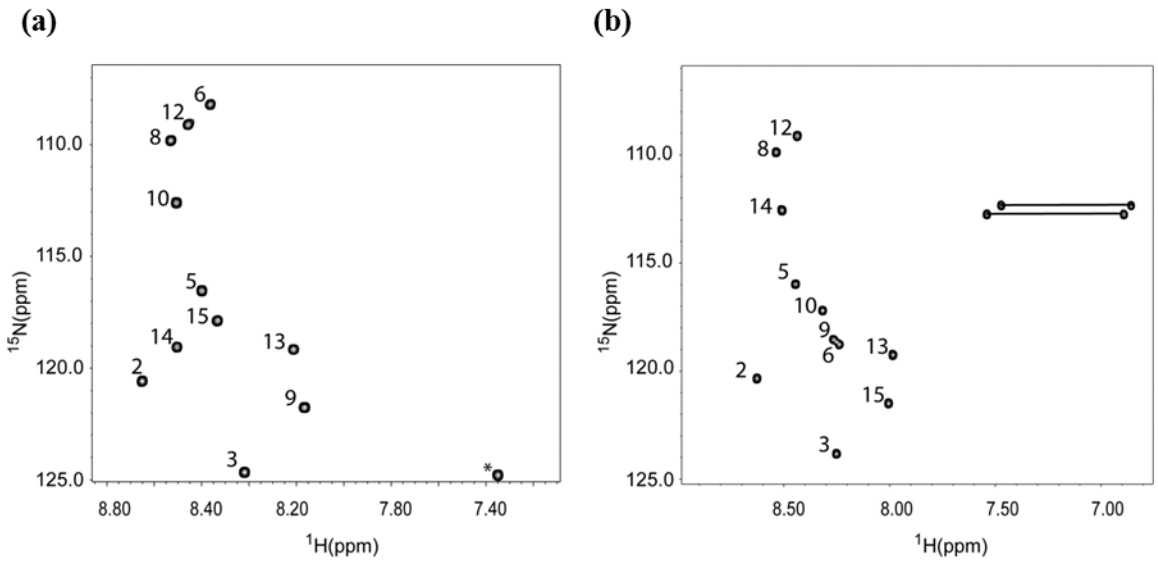


Figure 2. NHSQC spectra of MhCEP11 and MtCEP1.

The ^{1}H-^{15}N Heteronuclear single quantum coherence (HSQC) spectra for the MhCEP11 and MtCEP1 peptides are shown, respectively. The H-N peaks are labeled with the assigned residues for clarity. Side chain NH2's of MtCEP1 as denoted with a line connecting them in panel (b). "*" denotes the possible side-chain HN of one of the Histidine residues in MhCEP11 in panel (a).

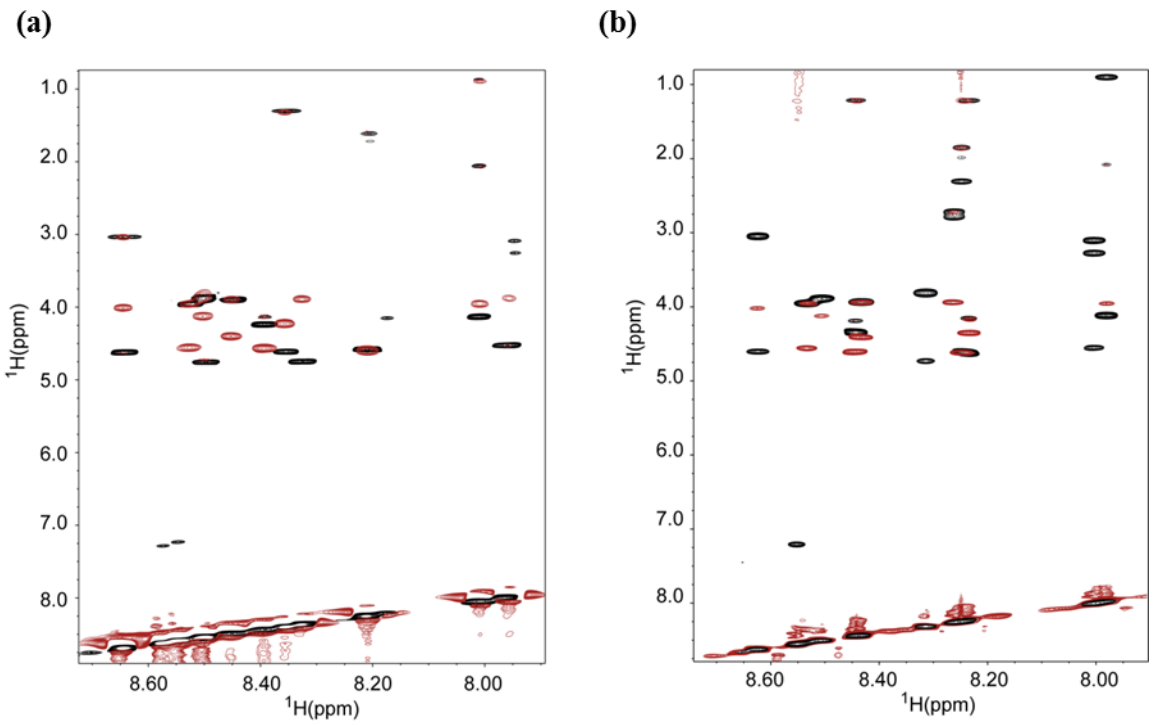


Figure 3. Total correlation and Nuclear Overhauser Effect spectra.

(a) MhCEP11 peptide's total correlation spectrum collected at 80 msec (black) overlaid with a Nuclear Overhauser effect spectrum collected at 300 msec (red).

(b) MtCEP1 peptide's total correlation spectrum collected at 80 msec (black) overlaid with a Nuclear Overhauser effect spectrum collected at 300 msec (red). Inter-residue through space connections contributing to the fold and structure of the peptides are shown as differences in the spectra (within panel a and b, themselves) between black and red peaks (where the peaks do not overlay on top of one another). Only the amide region is shown for clarity although the aliphatic region displayed similar characteristics.

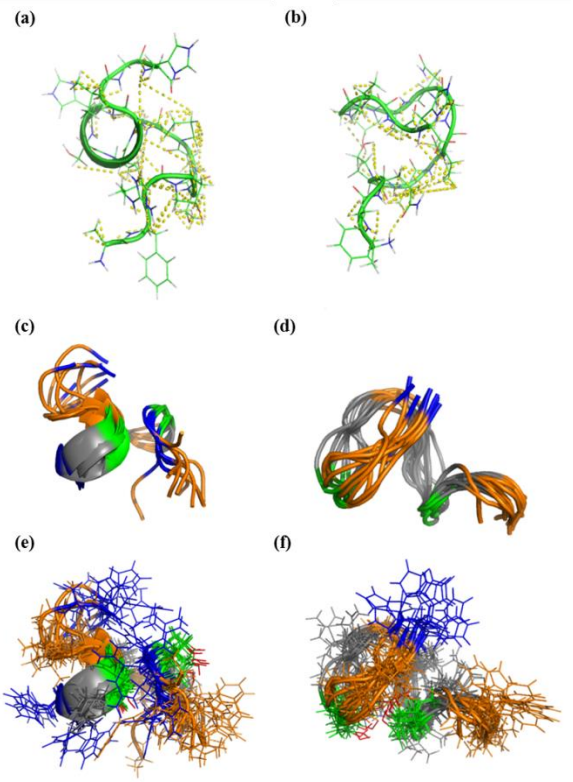


Figure 4. Solution structures of MhCEP11 and MtCEP1.

(a and b) represent the lowest energy structures in cartoon format and depict the NOE constraints for MhCEP11 and MtCEP1, respectively. The alpha helical character of the backbone is shown in cartoon format while all atoms (main chain and side chain) are shown in stick format. The yellow dashed lines denote the NOE connectivity (unambiguous and most highly populated ambiguous assignment – Table 1). (c - f) display the chemical characteristics of the peptide's structures; hydrophobic residues are colored orange, hydroxy-proline residues are colored green and positively charged residues of Arginine and Histidine are colored blue. The 10 lowest energy structures for MhCEP11 (c), and MtCEP1 (d), respectively. Panels E and F show the same orientation as panels (c) and (d) but with lines drawn for the individual amino acids to show dynamics of the residues' side-chains.

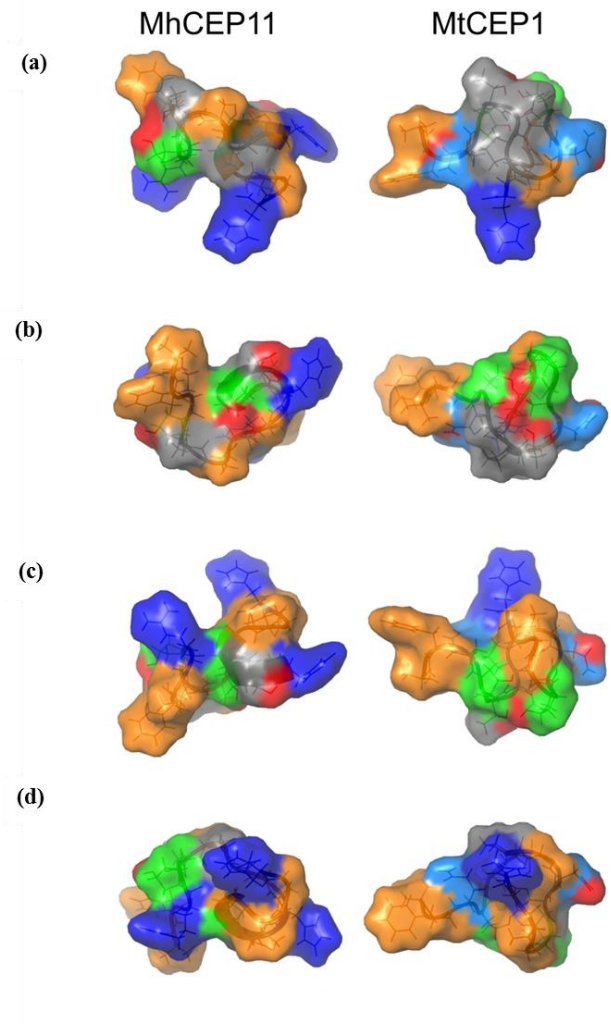


Figure 5. Surface characteristics of the lowest energy structure of MhCEP11 (left) and MtCEP1 (right).

(a - d) depict the peptide's chemical characteristics surface format; hydrophobic residues are colored orange, positively charged residues are colored blue, asparaginine residues are colored light blue, proline residues are colored green and the hydroxylated site on the proline is colored red. The peptide structures were aligned to one another based on peptide sequence and the panels are consecutive 90 degree rotations.

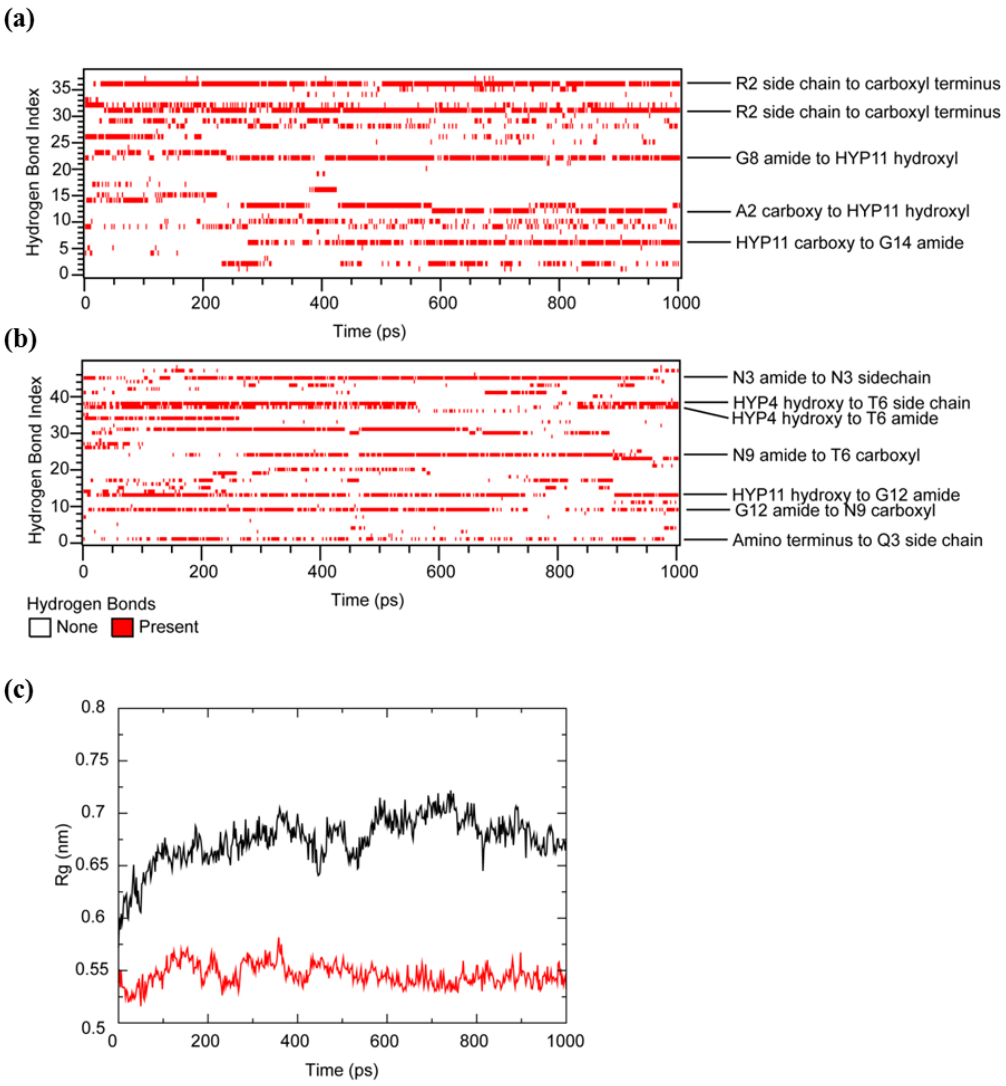


Figure 6. Molecular dynamics simulation.

(a ,b) depict hydrogen bond patterns for MhCEP11 and MtCEP1 during the MD simulation, respectively. The Hydrogen Bond Index (y-axis) represents a particular hydrogen bond formed. The highest populated hydrogen bonds observed are denoted to the side of the figures for each peptide. (c) displays the radius of gyration of MhCEP11 (red) and MtCEP1 (black) and is shown as a function of the time in the MD simulation. The radius of gyration is indicative of the level of compaction in the structure.

CHAPTER 3

Structural comparisons of host plant CLE peptides and parasitic root-knot nematode encoded mimics

ABSTRACT

The plant mechanisms appropriated by the parasite root-knot nematode (RKN; *Meloidogyne* spp.) to establish feeding sites from host cells, and ultimately, to reduce crop yield, remain largely unknown. Concurrently, identifying the signals and pathways responsible for normal plant cell development is encumbered by large, functionally-redundant families of regulatory molecules. Here I demonstrate that the RKN-host interaction offers a unique perspective, and can be used to characterize the roles of endogenous signals and pathways. I hypothesized that the identity of the RKN signals involved in communication with the host would mirror endogenous signals; our mining of the completed RKN genomes revealed the occurrence of 7 plant *CLE*-like genes. In plants, CLE are peptide hormones that participate in many aspects of plant cell development. Although there is an abundance of data on endogenous CLE activity, the function of the small, typically 12 amino acid ligands, cannot be determined solely by sequence or bioassay. Comprehensive expression analysis validated our gene predictions and *in vitro* studies demonstrated the functional equivalence of RKN-encoded CLEs with the endogenous ligands. To augment our genomic and functional analyses, I solved and compared the tertiary structures of host (*Arabidopsis*) and RKN encoded CLE peptides that have discrete functions. Molecular dynamics simulations and receptor docking models supplemented our structural and biological analyses; further supporting roles for

RKN encoded CLE mimics in the host-parasite interaction. These findings not only informs roles for endogenous and RKN encoded CLEs, but provides evidence for the integration of RKN parasitism with basic plant biology.

INTRODUCTION

The genus *Meloidogyne* (root-knot nematode; RKN) probably infects all seed plant species; this causes a reduction in the yield of infected crops equating to an annual economic impact approaching USD100 billion worldwide (Chitwood, 2003). Although multiple RKN gene products implicated in plant parasitism have been identified (Hewezi and Baum, 2013), the molecular basis behind nematode-plant communication remains ill-defined. As an obligate parasite, RKN biology likely mirrors aspects of host biology, which can inform the roles of endogenous pathways intrinsic to both host and parasitic biology. Importantly, all life stages of the nematode within host plants occur within the apoplast; RKN migrates intercellularly into the vasculature tissue of the host where copious amounts of protein, including cell-wall degrading enzymes, are visibly secreted from the feeding stylet (Davis and Mitchum, 2005). Within the apoplast of the vascular cylinder, RKN L2 manipulate up to ten vascular parenchyma cells which subsequently undergo developmental reprogramming into a unique cell type termed a giant cell (GC), which provide the nematode with nutrients. One model (Bird, 1996) postulates that GC are a novel chimera of a partially differentiated xylem cell and a transfer cell. Transfer cells form in response to a metabolic sink, which is provided by the feeding nematode. RKN induced GC contain many polyploid nuclei, the result of multiple rounds of nuclear division without cytokinesis, and have thickened cell walls and an

increased number of organelles. Depending on the RKN species or isolate, tissue surrounding GC undergoes variable degrees of hyperplasia, creating the noticeable galls (knots), which are the distinctive root phenotype of RKN infection. The formation both of GC and the gall suggest multiple plant cell developmental pathways are involved in RKN feeding site formation.

Linford's hypothesis (1937) that nematode secretions are involved in the parasitic interaction remains relevant today, and as recently reviewed (Hewezi and Baum, 2013), the identity of numerous individual proteins in RKN pharyngeal glands and secretions has been established (Ding et al., 1998; Smant et al., 1998; de Boer et al., 1999; Rosso et al., 1999; Gao et al., 2002; Qin et al., 2004; Huang et al., 2006a; Bellafiore et al., 2008; Xue et al., 2013). Manipulation of plant cell development by RKN secretions is certainly constrained by the repertoire of competent pathways available in the host. Much literature points to plant cell growth and development being largely specified by the interplay of the classical secondary metabolic hormones; and there is direct evidence that RKN and cyst nematodes (CN; *Heterodera* spp., *Globodera* spp.) influence the levels of these phytohormones (reviewed by Goverse and Bird, 2011) although the specific mechanisms are generally unknown. Over the past decade however, peptide hormones have been recognized as significant plant cell development regulators. Whole genome analysis has revealed that vascular plants also encode large numbers of trans-membrane, receptor-like kinases (RLK) and receptor-like proteins (RLP) that represent potential targets for peptide hormone ligands. Arabidopsis encodes at least 235 leucine-rich repeat RLK, including such members such as CLAVATA-

like element receptor and TDIF receptor (Sharma et al., 2003; Hirakawa et al., 2008; ten Hove et al., 2011). Collectively, this implies numerous opportunities for peptide signaling to regulate many diverse aspects of plant development, including parasitic nematode feeding sites. Indeed, included in the secretion catalogs for cyst nematodes were peptides similar to the CLAVATA-like element (CLE) class of hormones (Wang et al., 2001; Gao et al., 2001, 2003; Huang et al., 2003). Functional analysis has shown that the cyst nematode-encoded CLE hormone mimic(s) are biologically active in plants, rescue CLE knockout mutation phenotypes (Wang et al., 2005; Huang et al., 2006b; Mitchum et al., 2008) and appear to play key roles in establishing the feeding site (Replogle et al., 2013).

Like the CN CLE peptide mimics, an *M. incognita* protein designated 16D10 was initially isolated as a secreted protein (Huang et al., 2003) and later computationally identified as having sequence similarity to the CLE motif (Huang et al., 2006a). Transgenic over-expression *in planta* of 16D10 gave a root developmental response. However, it was found that the nematode ligand bound to two host SCARECROW-LIKE (SCL) proteins, in part from a yeast 2-hybrid assay (Huang et al., 2006a). SCL are members of the GRAS class of transcription regulators, which play central roles in root meristem specification and also are central to rhizobial nodulation (Hirsch et al., 2009), which is a process with many molecular and developmental similarities to RKN giant cell induction (Bird, 2004; Weerasinghe et al., 2005). The finding that 16D10 binds cytoplasmic SCL, and not an extracellular receptor has led to its dismissal as a CLE mimic (Mitchum et al., 2008).

In Arabidopsis, the CLE ligand hormone family constitutes 32 members (Kinoshita et al., 2008). Given the number of CLEs and also the large number of transmembrane receptors in plants, opportunities for CLE-type regulation are large. Adding to the complexity of this system, the activation of signal cascades, primarily understood as kinase cascades, can differ depending on the receptor a single CLE ligand binds (Betsuyaku et al., 2010). Recent developments in the understanding of native plant peptide hormone action have illuminated possible endogenous mechanisms exploited by plant parasitic nematodes including RKN and CN. Functional analyses of plant CLE have defined multiple classes. A-type CLE, which includes *Arabidopsis* CLE10, promote cell differentiation at the meristem by antagonizing the transcription factor WUS and aborting root growth. B-type CLEs, which includes *Arabidopsis* CLE44, do not promote cell differentiation at the meristem, but rather inhibit cell differentiation in *Zinnia elegans* tracheary elements. The action of the two CLE classes are not necessarily in opposition, more accurately, they are described as being agonistic, where A-type potentiate the activity of B-type CLEs (Whitford et al., 2008). This is demonstrated within the development of the vascular system of plants as the phloem and xylem are separated by the procambium/cambium which gives rise to both of these tissues on either side. To form this complex system, there must be a degree of communication able to balance proliferation with differentiation. Similar balances may be functioning to initiate and maintain RKN and CN feeding sites. This is especially apparent with RKN GC which may be partially differentiated xylem progenitors (Bird, 1996).

Placing the role of a single nematode secreted protein into the context of parasitism is hampered by many genus-specific technical barriers, especially when considering peptides with high sequence identity between parasite and host. Importantly, controlling for possible cross reactivity when using monoclonal antibodies to localize CLE peptide function *in planta* is confounded by limited number of available epitopes for a dodeca peptide; similarly designing specific RNAi constructs is restricted by sequence parology in large multigene families. Currently, data indicate the presence of SCN secretions within host cell cytoplasm (Wang et al., 2010); it is important to stress that no such data exists for any RKN effectors, although there are published reports that RKN secreted proteins localize to GC nuclei (Jaounnet et al., 2012; Lin et al, 2012). Recently, the variable domain within CN CLE (*Heterodera glycines;* HgCLE) has been implicated in the transportation of HgCLE from the host cytoplasm to the apoplast where it may act through innate pathways (Wang et al., 2010). However, GFP-tagged antibodies raised against HgCLE possibly indicate localization to syncytia cytoplasm. In the same report, evidence derived from the transient over-expression of protein fusions in plants and subsequent bioassays indicates an apoplastic mode of action. Fusion proteins were constructed by switching the variable domains of native CLEs and HgCLEs. Based on phenotypic analysis of root development, the variable domains between HgCLE and native CLEs are reported as interchangeable, despite the contradicting ascribed functions (Wang et al., 2010).

Plant parasitic nematode encoded CLE peptide mimics have been previously described with demonstrated expression and secretion during the parasitic interaction. Here, we describe

RKN CLE peptide mimics and augment these previous studies with well-defined NMR deduced tertiary structures. Further, we model RKN CLE mimic binding with receptors against genetically established plant CLE-receptor pairs. The relevance of CLE function to RKN feeding site development and pathology is discussed as well as the implications these findings have to plant biology.

MATERIALS AND METHODS

RKN CLE gene expression: Medicago truncatula cv. Jemalong A17 was inoculated with *Meloidogyne hapla* second stage juveniles. Three weeks post infection, RKN-induced galls were isolated and immediately frozen in liquid nitrogen. Total RNA was isolated from *M. hapla* and *M. truncatula* tissue using the RNeasy Mini Kit (Qiagen). Messenger RNA was processed for sequencing using standard Illumina protocols. Following similar protocols, RNA was also collected from *M. hapla* egg, J2 and whole root tissue of *M. truncatula* containing RKN 1, 2, 4, 5, and 7 days post inoculation with J2. TopHat (Trapnell et al., 2009) and Bowtie (Langmead et al., 2009) were used to align reads to the *M. hapla* genome. Expression of *M. hapla* genes is illustrated as total read counts

Peptide synthesis for NMR and bioassays: The active 12-14 amino acid ligand domains for AtCLE44 (HEVPSGPNPISN); AtCLE10 (RLVPSGPNPLHN); MhCLE4 (HEVPSGPNPSSN); MhCLE5 (RKVPTGSNPQKN) and MhCLE6/7 (HQVPSGPNPLHNKK) were deduced from the appropriate genome sequence, and synthesized to >98% purity, as validated by mass-spectroscopy and HPLC. Syntheses and

purity analysis were performed by ChiScientific (Boston, MA, USA). Peptide function was assayed on *Arabidopsis thaliana* plants using standard methods (Ohyama et al., 2008).

NMR spectroscopy: NMR spectroscopy follow established protocols detailed in Chpater 2 of this work. Briefly, a 16.4 T spectrometer at controlled temperature (298.15 ± 0.1 K) on a Bruker Avance 700 MHz equipped with a TCI cryogenic probe was used for all structure experiments. Peptide samples (4 mg/ml) were dissolved in 90% H_2O, 10% D_2O (v/v). DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) was used as an internal reference standard. Standard pulse sequences were recorded with two-dimensional natural abundance ^{15}N-heteronuclear single-quantum coherence (HSQC), ^{13}C-HSQC, ^{1}H-^{1}H total correlation spectroscopy (TOCSY) (Bax, 1969) and ^{1}H-^{1}H nuclear Overhauser effect spectroscopy (NOESY) (Wüthrich, 1986) spectra. The mixing times for ^{1}H-^{1}H TOCSY were 30 and 80 ms and the ^{1}H-^{1}H NOESY spectra were collected with mixing times of 50 and 300 ms with a 3.0 s recycling delay. In all experiments, the water peak was suppressed by presaturation. NMRPipe (Delaglio et al., 1995) was used to process all of the NMR spectra, which were then analyzed with NMRViewJ (Johnson, 2004).

Peptide structure determination: NMRViewJ software was used to integrate NOE cross-peaks (Johnson, 2004). ARIA (Ambiguous Restraints for Iterative Assignment - version 2.3; under default parameters) was used to perform structure calculations (Habeck et al., 2004; Linge et al., 2003). The 20 lowest energy structures from each iteration were used for NOE assignment and used for each subsequent iteration. After eight iterations, the 20 lowest

energy structures were refined by a molecular dynamics step in explicit solvent (water). From this, the 10 structures with the lowest energy were considered as accurately characterizing the peptide structures and are presented. Representation and quantitative analysis of the calculated structures were performed using PyMOL (Version 1.5.0.4 Schrödinger, LLC).

Molecular dynamics simulation: The MD simulations were performed with the GROMACS 4.4.5 software package using the AMBER 99sb-ildn force field and the flexible SPC water model. The initial structure (average energy minimized structure from the ARIA output anaylsis) was immersed in a periodic water box of cube shape (1 nm thickness) and neutralized with Cl- counterions. Electrostatic energy was calculated using the particle mesh Ewald method. Cutoff distances for the calculation of the Coulomb and van der Waals interaction were 1.0 and 1.4 nm, respectively. After energy minimization using a steepest decent method, the system was subject to equilibration at 300 K and normal pressure for 100 pS under the conditions of position restraints for heavy atoms and LINCS constraints for all bonds. The system was coupled to the external bath by the Parrinello-Rahman pressure and temperature coupling. The final MD calculations were performed under the same conditions except that the position restraints were removed and the simulation was run for 50 nS. The results were analyzed using the standard software provided by the GROMACS package.

Three-dimensional structure prediction and dock modeling of CLE receptors: The extra-cellular domain of *Arabidopsis* CLV2 (At1g65380) and TDR (At5g61480) were computationally identified using SignalP-HMM (Nielsen and Krogh, 1998) and TMHMM

(Krogh et al., 2001) to exclude secretion signals and transmembrane domains, respectively from the ecto-domain containing ligand binding sites. Three-dimensional protein structures and docking models were predicted as described by Meng and Feldman (2010). Briefly, homology modeling was used to construct the three-dimensional model of CLV2 and TDR using template amino acid sequences of plant steroid receptors for which X-ray crystal structures are available. Template sequences homologous to the CLV2 and TDR sequences were retrieved using a BLAST search in the PDB. The X-ray crystal structure of the plant steroid receptor BRI1 ectodomain was identified as the best choice based on sequence BLAST of the PDB for homology modeling (PDB code: 3RIZ) (PubMed**:** 21666665). An amino acid multiple sequence alignment of CLV2 or TDR and the BRI1 ectodomain template was generated using ClustalW (PMID: 12824354 and 10964570). CLV2 and TDR homology models were generated using Modeller 9 version 7 (Eswar et al., 2006). Stereochemical analyses of the refined models were carried out using Ramachandran plots by MolProbity website (Chen et al., 2010). Dock modeling of CLE ligands with CLV2 and TDR receptors was conducted using PatchDock and refined with FireDock (Schneidman-Duhovny et al. 2005; Mashiach et al. 2008).

RESULTS

Root-knot nematode express bioactive mimics of CLE peptide hormones: Our computational screen revealed 7 genes throughout the *M. hapla* genome with coding potential for secreted CLE-like peptides, summarized in the LogoPlots (Figure 1). The available genomes of migratory plant parasitic nematodes (*Radopholus similis* and *Pratylenchus coffeae*) or the

free living nematode *C. elegans* were also screened and in no instance were CLE-like genes detected. The *M. hapla* CLE-like genes are dispersed throughout the genome without any obvious grouping and were numbered by the order of the spanning contig; lower numbered *MhCLEs* are contained within lower numbered contigs. *MhCLE6* and *MhCLE7* are identical at the nucleotide level through the entire coding region, rendering them indistinguishable by expression studies and bioassays; therefor these genes were annotated as *MhCLE6/7*. All seven *MhCLE* genes are encoded as single exons with an average size of 145 nucleotides. Examination of *M. hapla* CLE sequences by SignalP4.0 revealed well-defined secretion signal sequences with a cleavage site directly amino to the canonical CLE ligand domain. In most cases (excluding MhCLE1 and MhCLE2), *M. hapla* CLE-like proteins terminate directly after the conserved CLE ligand domain and do not contain any additional amino acids on the carboxyl-terminus (Figure S1).

To establish the validity of the computationally identified RKN *CLE*-like genes, we conducted two complimentary RNA-Seq experiments that sampled the *M. hapla* transcriptome at various time points reflecting multiple stages of RKN parasitism. First, 115 independent biological replicates were sequenced representing *M. hapla* induced knots three weeks post infection with L2. This experiment yielded over 4.4 billion high quality Illumina generated reads that could be mapped to the complete *M. hapla* genome or host *Medicago truncatula*. This data set revealed expression of all 7 *M. hapla* CLEs, with drastically variable expression levels. The genes most highly expressed, *MhCLE6/7*, accounted for almost 88% of all *MhCLE* mapped reads. Of interest to this research, *MhCLE4* and *MhCLE5*

mapped reads accounted for 7.8% and 0.17% of all *MhCLE* reads. In addition to this data set, a longitudinal RNA-Seq experiment confirmed *MhCLE* expression in egg and pre-infective J2. *MhCLE4* was expressed at 2.64E-05% and7.36E-05% of all EST reads for egg and L2, respectively. These expression analyses were not devised to determine differential expression; rather, they were used to confirm the expression of *M. hapla* CLE genes during the parasitic interaction.

Based on a BLAST analysis of MhCLE ligand domains against *Arabidopsis* CLE ligands, it was possible to bin *M. hapla* encoded CLEs into multiple classes, including *Arabidopsis* type-A CLEs (e.g. AtCLV3, AtCLE9/10) and type-B CLEs (e.g. AtCLE41/44). A standard root growth bioassay was used to determine RKN CLE bioactivity and to distinguish and validate RKN CLE functional equivalence. As shown in Figure 1, *Arabidopsis* seedlings respond to exogenously applied CLE peptides whether the sequence is from an *Arabidopsis* gene or a RKN gene. Synthetic *M. hapla* CLEs predicted to be type-A based on primary structure, elicit the anticipated phenotype of stunted root growth (MhCLE3 and MhCLE6/7); and those predicted to be *M. hapla* B-type CLEs (MhCLEB4) do not, with one exception (MhCLE5). Importantly, the magnitude of the change is the same as the *Arabidopsis* control. MhCLE5 has highest sequence similarity to the A-type *Arabidopsis* CLE20 (At1g05065) yet did not elicit the expected root stunting in our assay.

Plant encoded CLE peptides and RKN mimics have defined structures: To augment our genomic and bioactivity I used NMR to determine the tertiary structures of RKN encoded and endogenous CLE peptides (Figure 2). Previous reports detailing endogenous CLE ligand biochemistry allows for comparative analysis between host CLE and RKN CLE mimics. I focused on MhCLE4, MhCLE5, MhCLE6/7 and *Arabidopsis thaliana* encoded CLE44 (AtCLE44; At4g13195) and CLE10 (AtCLE10; At1g69320) based on expression studies and activity in bioassays. The sequence similarity between endogenous and RKN encoded CLE is most apparent within the core, residues 3-9, with the sequence of 3-VP(S/T)G(P/S)NP-9 is seen in all peptides for which three dimensional structures were determined. The core is composed of 5 (out of 7) identical residues at positions 3, 4, 6, 8 and 9. Positions 5 and 7 are strongly (S vs T) and weakly (P vs S) similar, respectively (Figure 3). The NMR determined structures for the carboxyl-terminal active domains of and *Meloidogyne hapla* encoded CLE mimics MhCLE4, MhCLE5, and MhCLE6/7 as well as AtCLE10 and AtCLE44 are well defined with good structural qualities for folded peptides (Figure 2, Table 1). Well-defined structures are structures with high confidence and restraints that distinguish them from un-folded proteins. The Cα root-mean-square deviation (r.m.s.d) values for all residues on the backbone for AtCLE44, AtCLE10, MhCLE4, MhCLE5 and MhCLE6/7 are: 1.05Å, 1.88Å, 1.24Å, 2.09Å and 0.92Å respectively. The Cα r.m.s.d values for all heavy atoms for AtCLE44, AtCLE10, MhCLE4, MhCLE5 and MhCLE6/7 are; 1.85Å, 2.83Å, 2.11Å, 3.49Å, and 1.84Å respectively. Comparably larger heavy-atom r.m.s.d values indicate that the side-chains are more dynamic and sample a larger conformational space compared to the backbone atoms. The MhCLE5 structure is the most dynamic with backbone Cα r.m.s.d

value of 2.09Å and all heavy atom value of 3.49Å. When the backbone r.m.s.d for MhCLE5 is calculated using only residues 1-9, the value is reduced to 2.07 ± 0.29Å, which is more comparable to the other peptides.

Ramachandran analysis for all the peptides revealed 90% or greater of the *phi*/*psi* angles within generously allowed space or better. Two structures, AtCLE44 and MhCLEB4 reveal the best Ramachandran space for phi/psi angles with no residues in disallowed space. AtCLE10 has the greatest amount of disallowed residues in Ramachandran space with 10%. The residues that are in this space are Val3 and Leu10, each contributing 3 residues in disallowed space out of 10 possibilities each. The residues for all peptide structures that are in disallowed space are located at the beginning and end of the core structure (Figure 2) and in the termini.

Molecular dynamics simulations recapitulate CLE peptide classes: To understand the implications of the well-defined CLE structures, we simulated the molecular dynamics (MD) of each average energy minimized peptide structure from the output of ARIA in explicit water for 50 nano-seconds. All peptides reached equilibrium by 30 nano-seconds, for clarity only the first 30 nano-seconds are show in each figure. The r.m.s.d and radius of gyration (i.e. compactness of the structures) of the backbone of the peptides was monitored throughout the molecular dynamic simulation (Figure 3, 4). All peptides experienced an increase in the r.m.s.d compared to the structure at the start of the simulation indicative of conformational flexibility. Peptides AtCLE44 and MhCLEB4 reach equilibrium with regard

to their respective r.m.s.d values within the first 5 nano-seconds and average to a value of 0.44 and 0.52 nm. In contrast, AtCLE10, MhCLEB5 and MhCLE6/7 do not reach equilibrium until 20 nano-seconds and average 0.21, 0.71 and 0.68 nm respectively (Figure 4A). All peptides display an inherent flexibility of the backbone; however AtCLE10 displayed the most restricted backbone flexibility while MhCLE5 and MhCLE6/7 exhibited the greatest backbone flexibility.

The radius of gyration for each peptide was also monitored throughout the MD simulation. Peptides that maintain a constant radius of gyration throughout the simulation indicates a folded peptide maintaining its structure through a series of changing stable interactions which can include hydrogen bonding, salt bridges and van der Waals interactions. If a peptide were to become unfolded during the simulation, the radius of gyration would continue to increase throughout the experiment. All peptides retained a level of compactness over the course of the simulation, mainly imparted by the core residues (3-9), while the termini residues were more dynamic. However, not all peptides exhibited similar compactness (Figure 4B). AtCLE10 and MhCLE4 demonstrated similar and relatively greater stability with radius of gyrations at 0.59 nm and 0.57 nm. The radius of gyration for AtCLE44, MhCLE5 and MhCLE6/7 varied greatly, reaching equilibrium at 0.80 nm, 0.99 nm and 1.70 nm, respectively. However, as with the r.m.s.d, AtCLE44 and MhCLE4 reached equilibrium around 13 nano-seconds into the simulation while AtCLE10, MhCLE5 and MhCLE6/7 reach equilibrium together around 20 nano-seconds.

Arabidopsis CLE receptor modeling and docking of RKN and host CLE peptides reveal dynamic interaction: Using homology threading algorithms (Modeller; Eswar et al., 2006), I determined the tertiary structures of the extracellular domains of *Arabidopsis* receptor-like proteins (RLP) TDR (AT5G61480) and CLV2 (AT1G65380). Importantly, my model structures improve upon their previously modeled structures presented in Meng and Feldman (2010) due to additions to databases, which informs a more true structure. Both RLPs had the highest sequence homology and were modeled with the x-ray crystallography solved BRI1 steroid receptor (She et al., 2011; Hothron et al., 2011). *Arabidopsis* CLE44 (At4g13195) has been identified as interacting with AtTDR (Hirakawa et al., 2008), while CLE10 (At1g69320) has been shown to interact with AtCLV2 (Kondo et al., 2010). Using our NMR determined structures of *Arabidopsis* CLE and RKN encoded mimics, I modeled the binding interactions of CLE ligands with their respective receptors.

Our bioassays demonstrated that sequence similarity between RKN encoded CLE and *Arabidopsis* CLE ligands could not completely and determinably distinguish type-A from type-B CLEs. Here, we used docking models to further define the CLE ligand-receptor complex and resolve sequence and bioassay discrepancies (Figure 5). Using PatchDock and FireDock (Schneidman-Duhovny et al. 2005; Mashiach et al. 2008) to model ligand docking and refine the receptor complex, respectively, AtCLE44 was shown to bind AtTDR with a global binding free energy of -40.71. The negative free energy value is indicative of a favorable interaction. RKN encoded CLE4 (MhCLE4) has the highest sequence similarity to AtCLE44 and functions similarly in bioassays. When modeled with the AtTDR receptor,

MhCLE4 interacts in the same binding site and relatively similar free energy (-29.62) as its inate counterpart. AtCLE10 is known to interact with CLV2; this interaction was modeled with a global free energy of -7.37. MhCLE6/7 has the highest sequence similarity to AtCLE10 and is functionally equivalent in bioassays. When modeled with CLV2, again the RKN mimic binds the receptor at the same site and energy (-6.18) compared with the endogenous ligand. However, MhCLE5, has a sequence similar to endogenous A-type CLEs, yet did not inhibit root growth in bioassays. We modeled MhCLE5 docking with AtTDR and AtCLV2. MhCLE5 demonstrated drastically different binding energies (-6.15) and sites when modeled with AtTDR, a known B-type CLE receptor; but similar energies (-4.21) and binding sites when modeled with CLV2, a known A-type CLE receptor (Figure 5).

DISCUSSION

An independent line of evidence for the role of RKN-encoded CLE has been demonstrated by plant genetics; mutations in genes encoding CLE receptors exhibit RKN-related phenotypes. Specifically, *Har-1* encodes the *Lotus japonicus* ortholog of *Arabidopsis* CLV receptors and mutations in the *Har-1* gene result in plants that hyper-infect with RKN (Lohar and Bird, 2003). Here we have described the occurrence, expression, bioactivity and tertiary structures of CLE-like peptides encoded within RKN genomes, and discuss their relevance to RKN parasitism and plant biology.

In plants, intrinsic CLEs are secreted into the apoplast as pro-proteins. Cleavage of the pro-domain is necessary for activity, presumably serving as an additional regulatory check on the

activity of these potent ligands (Ni et al., 2011). In contrast, RKN CLEs are not predicted to encode a pro-domain. Rather, the signal sequence is immediately adjacent to the ligand domain. This likely permits RKN to directly secrete active peptides into the host apoplast, where, akin to intrinsic peptides, they are recognized by trans-membrane receptor kinases and initiate signal cascades, ultimately influencing plant cell development. A strikingly different model has been proposed for cyst-nematode CLE mimics. CN translate *CLE* mimics as pro-proteins, which, like endogenous CLE, contain a variable pro domain. It has been demonstrated that this variable domain is not only required for CN CLE function *in vivo*, but directs CN CLEs through host secretion pathways where they presumably bind transmembrane receptors in the host apoplast. CN CLEs are initially secreted by the nematode into host cytoplasm (Wang et al., 2010). Unlike RKN, cyst nematodes have a more restricted host range, but one that includes many important crops. In the U.S., soybean cyst nematode (*Heterodera glycines*) is a major problem. Although the feeding nematode also resides in the apoplast, direct evidence shows that they have access to plant cytoplasm into which they can inject proteins (Elling et al., 2007), and from which they presumably feed. Although deceptively similar in appearance, cyst nematode feeding sites are quite different from those induced by RKN. GC induced by RKN undergo multiple rounds of nuclear division without cytokinesis, cyst nematode J2 induce formation of a syncytium via the coalescence of numerous adjacent cells. Although once believed to be closely related, recent phylogenetic analyses (Holterman et al., 2008) show an ancient divergence for the ancestors of cyst and root-knot nematodes. These differences caution against assuming that the host parasite interactions between RKN and the plant will be mechanistically similar to those

involving cyst nematodes (including the role of CLE). On the other hand, because both genera pervert normal plant development, host biology presumably places constraints on the molecular nature of the interactions, demonstrated by the presence of CLE peptide mimics in these genera.

Deducing the role of CLE in plant parasitic nematode interactions is not straightforward. Over-expression analyses and bioassays are fraught with complications (Strabala, 2008); these include, but are in no terms limited to pleiotropic (and often opposing) phenotypes. This may be due to neomorphy, hypermorphy or antimorphy, a result of peptide concentrations and ectopic applications pushing the ligands kinetics outside normal physiological relevance (Strabala, 2008). These incongruences were observed in our bioassays, which could not completely distinguish type-A from type-B CLEs alone. Specifically, MhCLE5, which is most similar in sequence to the A-type *Arabidopsis* CLE20 did not inhibit root growth as expected. However, we are able to exploit this unique host-parasite relationship, to better understand CLE function. Since RKN encodes and expresses mimics of host proteins, comparative peptide structural analyses are able to augment expression analysis and bioassays. In addition, there is an abundance of functional and genomic data on host CLE peptides from which to direct parasite-encoded CLE functional investigations and which can inform CLE function in both RKN and host biology.

We used NMR to deduce the tertiary structures of *Arabidopsis* and *M. hapla* encoded CLE peptides. Plant CLE peptide ligands are extensively modified post-transnationally

(Matsubayashi, 2011), however, post-translational modifications are not required for CLE activity and were omitted from our synthetic peptides (Meng et al., 2010). Aside from mediating receptor recognition, post-translational modifications on peptide hormones have been speculated to function in resistance to enzyme degradation, regulating transport, and ligand storage (Matsubayashi, 2011). Some MhCLEs contain additional residues beyond the conserved CLE domain, at the carboxyl-terminus. There is precedent that nematode encoded CLE peptides are proteolytically processes by the host, specifically truncating residues outside of the ligand domain (Guo et al., 2011). Unlike CN CLEs, carboxyl-terminal extensions are relatively rare and seen in only 2 of the *M. hapla* CLE proteins (MhCLE1, MhCLE2).

Our structural and molecular dynamics analyses of the RKN and *Arabidopsis* encoded CLE peptides revealed several unique aspects of their structural features, and potentially indicate how these structural features relate to function. First the peptides do have defined structure despite their small size (Figure 2), distinguishing them from unfolded random coils. Secondly this defined structure is dominated by a core of residues found in each peptide, residues 3-9 (3-VP(S/T)G(P/S)NP-9). This core of residues may be a general receptor-recognition structural motif found in this class of peptides that structurally defines this class of peptides; the less conserved terminal residues may act to target specific receptors, and alter binding kinetics. *Arabidopsis* CLV3 is produced as a 12 or 13 aa peptide and the binding affinity for a CLV2-CRN complex is greater as a 13 aa peptide (Meng and Feldman, 2010). Furthermore, from the solved structures shown in Figure 2, one can observe the potential for

the residues outside the core residues to be quite flexible and sample a larger conformational space. This is recapitulated in the molecular dynamics simulation which revealed that the peptides retain the structural conformation of the core residues throughout the simulation (Figure 3). However, the side chain residues and the termini of each peptide are quite dynamic and are unique for each peptide with regard to the volume of conformational space sampled.

The peptides also tend to cluster according to their behavior in solution. Peptides AtCLE10, MhCLE5 and MhCLE6/7 tend to have structures that tend to be a little more dynamic in solution and do not settle into an equilibrated conformational space until about 20 ns into the simulation (Figure 4 – red, blue, and orange, respectively). Interestingly, these peptides are all classified as type-A CLEs based on sequence and/or bioassays. Peptides AtCLE44 and MhCLE4 tend to achieve this equilibrated conformational space quite quickly in comparison (~5-10 ns into the simulation – Figure 4 – black and green, respectively), again agreeing with sequence similarity and function in bioassays as type-B CLEs. Our cumulative data suggests that peptides AtCLE10, MhCLE5 and MhCLE6/7 are linked not only structurally but functionally and that this function is distinct from AtCLE44 and MhCLE4 which are similar, both structurally and functionally to one another.

Because genetically defined plant CLE ligand-receptor interactions are known, we were able to model these interactions. Using homology threading algorithms, we deduced the tertiary structures of the extracellular domains of Arabidopsis receptor-like proteins TDR and CLV2.

Using PatchDock to model ligand docking and FireDock to refine the receptor complex, AtCLE44 binds AtTDR with a global free energy of -40.71. The negative free energy value is indicative of a favorable interaction. RKN encoded CLE4 (MhCLE4) has the highest sequence similarity to AtCLE44 and functions similarly in bioassays. When modeled with the AtTDR receptor, MhCLE4 interacts in the same binding pocket and with a similar free energy (-29.62) as its endogenous counterpart, consistent with this being a *bona fide* interaction; we will test this hypothesis in future work. As shown in Figure 5, we have repeated these analyses for a number of peptides and receptors. Collectively, we have demonstrated that defined tertiary structures of CLE peptides and receptor-binding modeling can clarify potential bioassay/over-expression discrepancies seen with members of this large, often functionally redundant, yet tissue specific, gene family.

Although the receptor AtTDR contains a cytoplasmic kinase domain to relay extracellular signals to the nucleus (presumably through a kinase cascade) that alters gene expression and ultimately cell development, CLV2 lacks any defined cytoplasmic domain. Instead, CLV2 consists of only a transmembrane anchor and extracellular receptor. For a ligand to conduct a signal through CLV2, additional proteins are required to form hetero-complexes with CLV2. Multiple signal transduction models for CLV2 have been proposed, often involving hetero-multiplexes of multiple RLK and cytoplasmic kinases, and may be dependent upon tissue type (Brand et al., 2000; Müller et al., 2008; Bleckmann et al., 2010; Guo et al., 2010; Durbak and Tax, 2011). One model in particular has independent lines of supporting evidence; CLV2 requires the cytoplasmic-restricted CRN protein for localization in the

plasma membrane and relay of extracellular signals (Müller et al., 2008; Bleckmann et al., 2010; Meng and Feldman, 2010). Additonally, AtCLE14 and AtCLE20 were shown to interact with a CLV2/CRN hetero-complex (Meng and Feldman, 2010); docking models of these interactions included CRN forming a hetero-complex with the extracellular domain of CLV2. We did not include a cytoplasmic kinase domain in our CLV2 docking models; rather our aim was to compare ligand binding of known receptor-peptide pairs with RKN peptide mimics. Although our analyses indicate possible RKN CLE-receptor interaction, the identification of the true receptors and possible cytoplasmic kinase domains involved in RKN CLE function remains a continual goal.

Table 1. Details and statistics for quality assessment of peptide structures.

	AtCLE44	AtCLE10	MhCLE4	MhCLE5	MhCLE6/7
Total NOES	78	53	52	58	67
Intra-residue	46	35	29	38	49
Inter-residue	32	18	23	20	18
Unambiguous	52	42	38	48	46
Ambiguous	26	11	14	10	21
Average violations / structure					
> 0.5Å	0	0	0	0	0
> 0.3Å	0	0	0	0.1	0.3
> 0.1Å	2.6	2.6	2.6	1.8	3.7
Energy (kcal mol-1)					
van der Waals	-74.34 ± 1.65	-78.53 ± 3.00	-65.96 ± 2.47	-76.44 ± 3.00	-102.48 ± 4.40
electrostatic	-275.14 ± 27.18	-291.13 ± 33.86	-341.03 ± 37.67	-326.32 ± 28.36	-367.44 ± 38.77
Deviations from idealized geometry					
Bond lengths (Å)	0.0043 ± 0.0002	0.0035 ± 0.0002	0.0042 ± 0.0003	0.0037 ± 0.0004	0.0039 ± 0.0002
Bond angles (°)	0.66 ± 0.05	0.55 ± 0.07	0.63 ± 0.05	0.54 ± 0.07	0.67 ± 0.04
Impropers (°)	1.53 ± 0.29	1.26 ± 0.17	1.71 ± 0.31	1.26 ± 0.25	1.74 ± 0.23
RMSD (Å)					
Backbone	1.05 ± 0.23	1.88 ± 0.30	1.24 ± 0.34	2.09 ± 0.40	0.92 ± 0.29
Heavy atoms	1.85 ± 0.32	2.83 ± 0.31	2.11 ± 0.36	3.49 ± 0.44	1.84 ± 0.50
Ramachandran analysis					
Core	16.7	33.3	60.0	38.6	31.2
Allowed	71.7	55.0	33.3	50.0	55.0
Generous	11.7	1.7	6.7	5.7	6.2
Disallowed	0	10.0	0	5.7	7.5

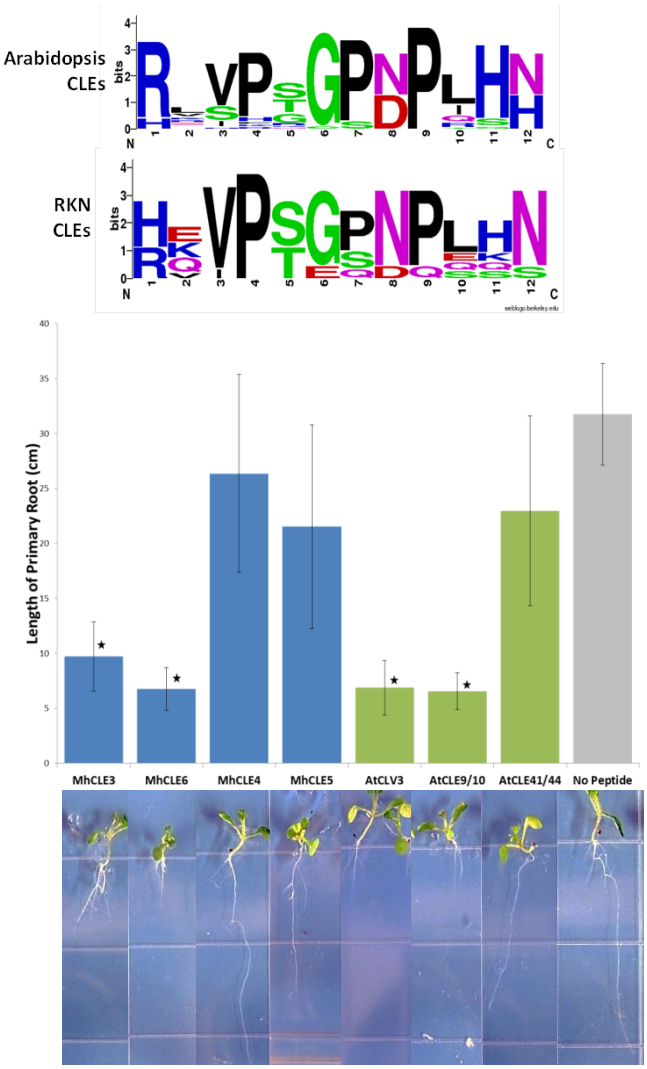


Figure 1. RKN encode bioactive mimics of CLE regulatory peptides.

(Top) Logo Plot of Arabidopsis and RKN CLE CLE active ligand domains show high sequence similarity, especially between the core residues 3 and 9. (Bottom) Standard root growth assays with exogenously applied synthetic peptides. (green, plant CLE ligands; blue, RKN CLE ligands).

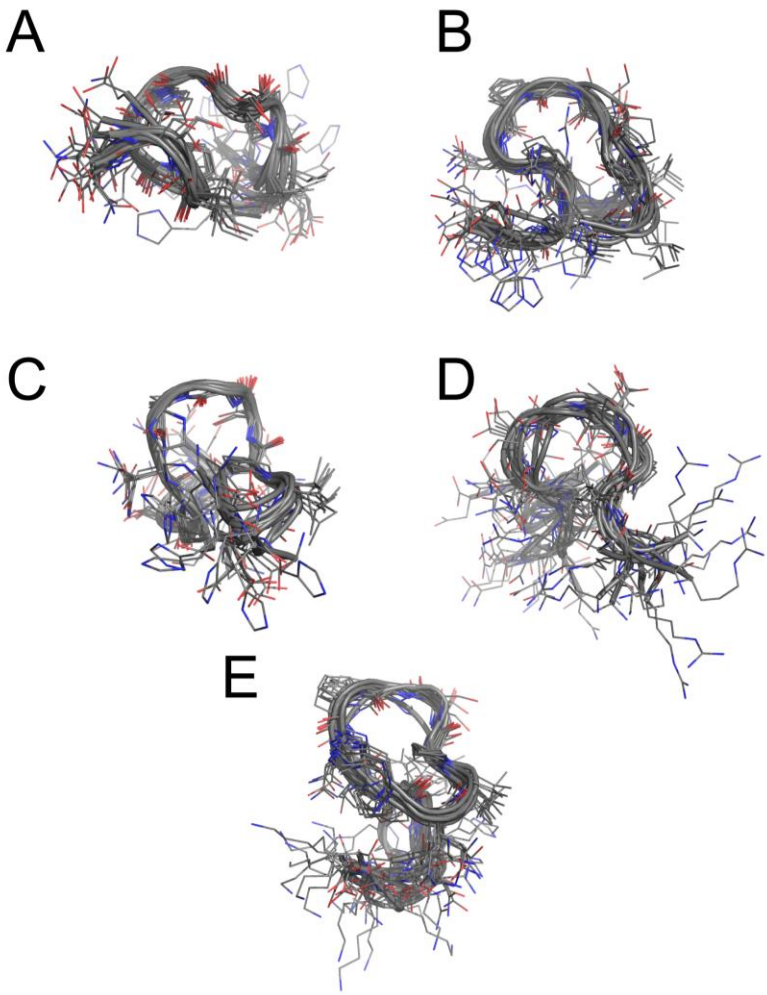


Figure 2. Structure overlay of the lowest 10 energy structures for plant and RKN CLE ligand domains.

(A) AtCLE44, (B) AtCLE10, (C) MhCLE4, (D) MhCLE5 and (E) MhCLE6. Structures are shown in cartoon format with residues shown in line format. All hydrogen atoms are removed for clarity, carbons are colored gray, oxygen atoms are colored red and nitrogen atoms are colored blue. Structures are aligned to the conserved sequence (residue 3-9) of VP(S/T)G(P/S)NP.

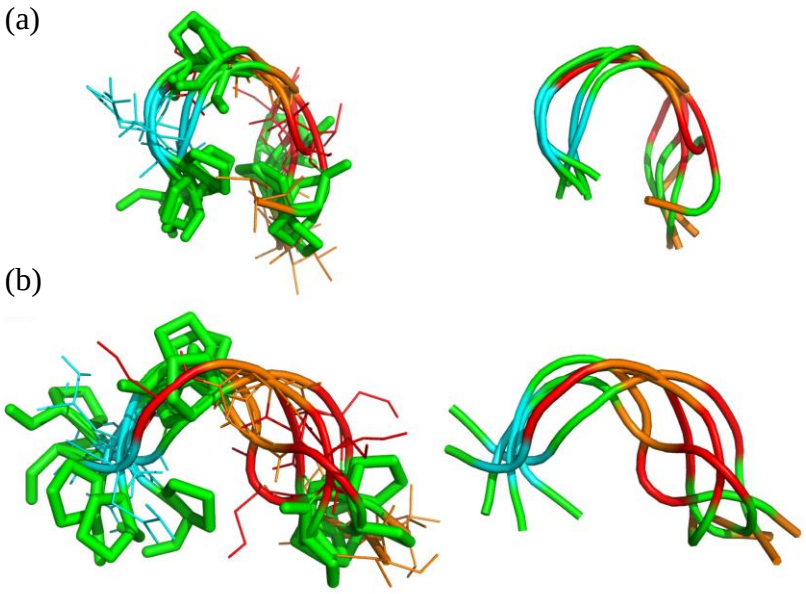


Figure 3. Plant and RKN share similar core structural features**.**

The peptide sequences have a core of residues, residue 3-9 that are highly conserved (VP(S/T)G(P/S)NP). This similar sequence is recapitulated as a similar structural motif between the peptides. (a) The average energy minimized structures for AtCLE44, AtCLE10, MhCLE4, MhCLE5 and MhCLE6/7 are overlaid with respect to residues 3-9. (b) The average structure from the highest populated cluster from the clustering analysis in gromacs based on a Cα r.m.s.d of 2Å. Shown on the left is the overlay with the proline residues colored green and shown in stick format while the hydrophobic residues (Gly and Val) are colored orange, Asn is colored cyan and Ser or Thr is colored red with line format drawn for clarity. Show on the right is the exact same thing with the sticks and lines removed for clarity.

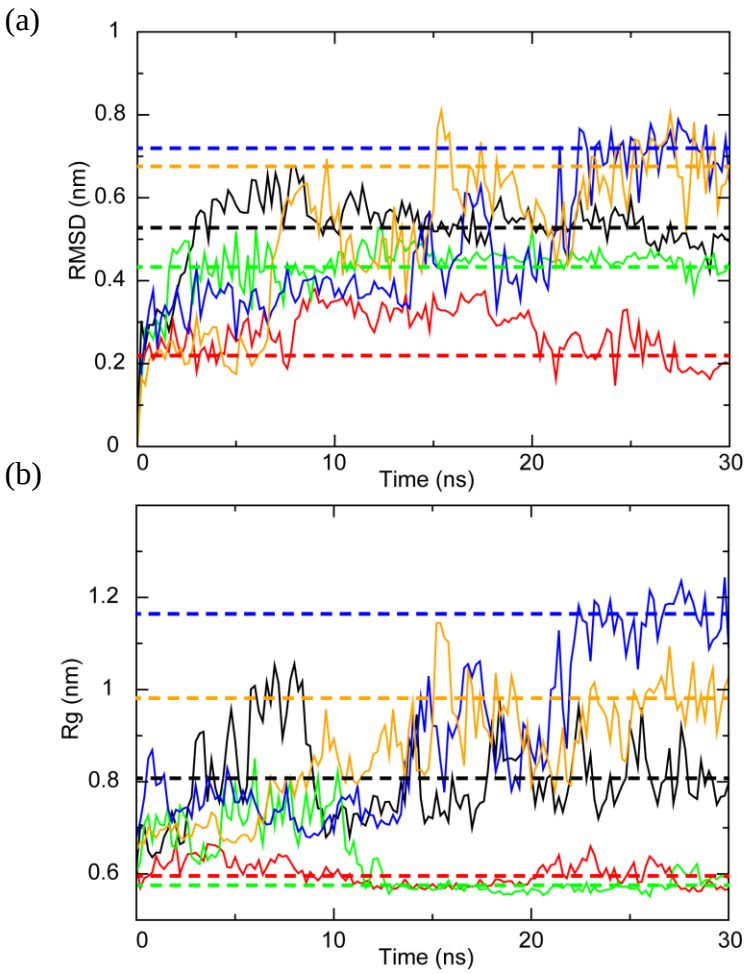


Figure 4. Molecular dynamics simulation of defined plant and RKN CLE structures.

(a) The backbone r.m.s.d of (black) AtCLE44, (red) AtCLE10, (green) MhCLE4, (blue) MhCLE5 and (orange) MhCLE6 is shown as a function of the time in the molecular dynamics simulation. The r.m.s.d is a measure of the peptide's conformational flexibility compared to the starting structure of the molecular dynamics simulation. (b) The radius of gyration of (black) AtCLE44, (red) AtCLE10, (green) MhCLE4, (blue) MhCLE5 and (orange) MhCLE6/7 is shown as a function of the time in the molecular dynamics simulation. The radius of gyration is indicative of the level of compaction in the structure. Colored dotted lines denote the final r.m.s.d and rg equilibrium points.

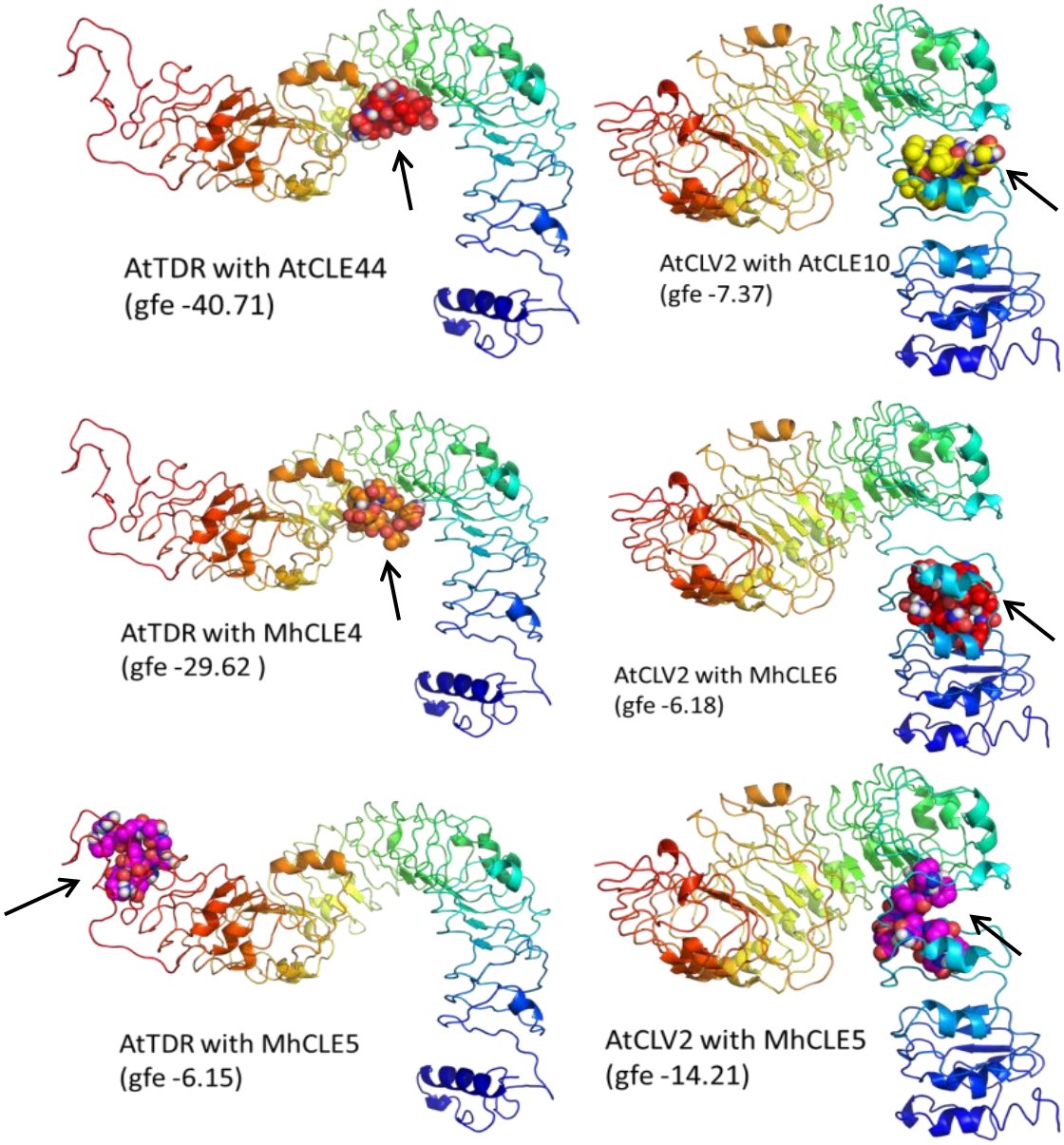


Figure 5. Ligand-Receptor complex docking models.

Known interactions (AtTDR with AtCLE44; AtCLV2 with AtCLE10; top right and left) are used as standards to compare possible RKN encoded CLE mimic receptor binding for both binding site and global free energy (gfe). Black arrows indicate peptide (ball and stick model) binding site. Receptors are shown as ribbon for clarity.

CHAPTER 4

Invasion of roots by Root-Knot Nematode is Diurnally Biased

ABSTRACT

Root-knot nematode (RKN; *Meloidogyne* spp.) parasitize virtually all crop species and cause extensive yield loss. Currently, there is a limited understanding of the intimacy between the parasitic nematode and its host. These obligate parasites are likely aware of changes in host biology and presumably respond accordingly. As many plant processes are regulated by, and dependent on, the hosts plants light-driven circadian clock, we hypothesized that RKN biology might be sensitive to these rhythms. Specifically, we investigated the influence plant diurnal rhythms have on RKN invasion of host roots using quantitative PCR. Multiple time points were examined, spanning the change from night to day using whole root tissue collected 24 hours post inoculation with naïve second stage *M. hapla* juveniles. Interestingly, we determined RKN invasion to be diurnally biased with significantly more nematodes invading host roots at night. The results of these assays shed light on RKN host perception and further demonstrate host biology as a critical arbitrator of nematode parasitism and emphasize the importance of considering host biology when studying plant parasitic nematodes.

INTRODUCTION

Plant parasitic nematodes inflict an enormous burden on world agriculture, with yield losses possibly exceeding USD100 billion annually (Chitwood, 2003). Root-knot nematodes (RKN: *Meloidogyne* spp) cause a significant amount of this economic damage by parasitizing roots. Given the substantial crop yield losses attributable to RKN, controlling these pests is a major agricultural priority. Current efforts focus largely on the development of general nematicides and post-invasion control in the form of resistance (*R*) genes. However, the availability of chemical pesticides is severely limited and host-resistance is not available for most crops. What remains crucially limiting is a thorough understanding of the mechanisms underpinning the nematode-plant interaction. While it is true that some components have been identified in the form of single gene effectors, these pathways are employed post penetration, after the nematode has hatched in the soil, and invaded a suitable host.

Here, we focus on pathways that may regulate RKN parasitism prior to host penetration. Multiple events throughout the nematode lifecycle are influenced by the perception of the host-, rhizosphere- and even nematode-derived signals. In particular, egg hatch, root-attraction and penetration each have been demonstrated to be, at least partially, influenced by external stimuli (Viglierchio and Lownsbery, 1960; Teillet et al., 2013). Novel control methods may be realized with an increased understanding of the signals influencing these events. Substantial interest and study has been placed on the signals regulating initial RKN host perception and penetration; RKN have demonstrated responsiveness to different chemical and molecular stimulants including salt ions, CO_2 and pH gradients (Riddle and

Bird, 1985; Wang et al., 2009). Yet the exact mechanism arbitrating nematode attraction to host roots has yet to be confirmed; for an extensive review see Curtis et al., (2009).

Plant biology is tightly linked to light, and is manifested as circadian and diurnal rhythms. Circadian rhythms refer to periodicity around a 24 hour clock while diurnal rhythms are directly linked to the presence or absence of light. Multiple plant metabolic, chemical, and physiological changes are regulated in these manners, including leaf starch metabolism and root growth (Yanzdanbakhsh et al., 2011), all of which can possibly influence nematode competence. Other pathogens have been observed to have a relationship dependent upon host's perception of light and diurnal rhythms such that pathogenicity is optimized for certain light-dependent host cues (Stone et al., 2012). There is also evidence that suggests nematodes near host roots send out aggregation signals to further attract other nematodes to sites of penetration (Wang et al., 2010). Here, we sought to determine if the plant signal(s) arbitrating nematode invasion are dependent on host diurnal rhythms. We used qPCR to quantify nematode penetration 24 hours post inoculation with naïve second stage juveniles. Within the 24 hour timeframe, the nematode would have had to perceive the host, migrate to the site of penetration, and successfully penetrate the host to be detected in our system. Although all involved signals are of great interest (nematode, rhizoshpere and host derived), here we focused on how changes in host biology, specifically alterations due to changes in light, alter nematode host perception and ultimately, penetration.

MATERIALS AND METHODS

Plant growth: Medicago truncatula cv. Jemalong A17 was used in all experiments. Seedlings were scarified in concentrated H_2SO_4 for 15 minutes or until dark spots appeared on the seeds, which were then washed in sterile water until all residual H_2SO_4 was removed. Seeds were germinated in artificial soil for one week before being transplanted into sterile sand. Fertilization was done three times a week with 0.5x Hoaglands solution (Hoagland and Arnon, 1950). Host plants were maintained in a growth chamber set at constant humidity and temperature of 24.5^{o} C with an 18 hr day cycle. Three weeks after germination, seedlings were inoculated with 1,000 second stage *M. hapla* VW9 juveniles (J2).

Nematodes: M. hapla lines used in all experiments were maintained on Tomato (cv. Rutgers) in greenhouse conditions. RKN eggs were isolated using the standard bleach method and purified on a 40% sucrose gradient as previously described (Byrd et al., 1966). Sterile eggs were hatched in 'hatcheries' consisting of a small bowl containing a mesh screen supporting multiple layers of lab tissue. Viable, hatched RKN J2 were collected after three days. RKN J2 were kept naïve in sterile water and were used in infect host plants to optimize synchronous infection.

Root staining: To confirm RKN J2 host penetration 24 hours post inoculation, roots were stained with acid fuchsin (Byrd et al., 1983). Briefly, isolated and washed root systems were cleared with a 1% NaOCl solution for 4 minutes and rinsed multiple times in distilled water. Acid fuchsin stain was prepared by dissolving 3.5 g of acid fuchsin in 250 mL of glacial

$C_2H_4O_2$ and diluting with 750 mL dH_2O. This stock was used at 2% to cleared roots in 50 mL. The cleared root and stain were heated to just before boiling and allowed to cool to room temperature before the roots were rinsed in tap water. The stained root systems were then placed in acidified glycerol for de-staining and to increase stain contrast. The acidified glycerol solution (few drops of 5M HCl in 30 mL glycerin) containing the root system was brought to almost a boil and after the roots cooled to room temperature they were examined under a dissecting microscope in acidified glycerol.

Nematode quantification using qPCR: The nematode sequence, referred to as MhC217 is located at on MhA1_contig217:72348..72400 and when queried against the non-redundant database returned no other significant BLAST results. The host sequence used for normalization and control, referred to as MtUBQ1, (MtGI accession number TC161574) also returned no other significant BLAST results when queried against the non-redundant nucleotide database. The sequences were also selected based on constraints imposed for optimal primer design, product length, GC content and use in SYBR Green qPCR reactions. Primer sequences MtUBQ1 Forward: gaacttgttgcatgggtcttga, Reverse: cattaagtttgacaaagagaaagagacaga; MhC217 Forward: atagttggtggcctctgtcc, Reverse: tggaatgtgctatttccaagggt). Quantitative PCR (qPCR) was conducted using the Applied Biosystems StepOne machine in 96-well plates containing 20µL reactions consisting of diluted sample DNA, 0.5µM of each primer and appropriate SYBR Green Master Mix. Time points were selected to cover the change from night to day; dark time points are represented by negative numbers as hours before daybreak and light time points are represented by

positive numbers as hours after daybreak (t = -2, +1, +3, +5). Eight experimental replicates were conducted with at least three plants each, yielding >24 whole root samples per time point. No-template controls were included to account for non-specific primer amplification.

Statistics: To quantitate qPCR results, the returned and averaged C_T values of MtUBQ1 for each time point was subtracted from each of the corresponding C_T values of MhC217 ($\Delta Ct = C_T^{MhC217} - C_T^{averaged\_MtUBQ10}$. This was done to account for, and to normalize any bias root mass may have on nematode penetration of host roots. The inverse of this computation was plotted to illustrate the representative quantification of nematode abundance in root systems (Figure 2). ANOVA (Table 1) and student t-tests were used on all six possible pairwise comparisons. Using the highly conservative Bonferroni multiple testing correction (SAS software), significance was determined as $p < 1.7E-2$; see summary in Table 2. Specifically, student t-test was conducted on the difference between normalized MhC217 C_T values ($\Delta\Delta C_T = \Delta C_T^{time\ point\ A} - \Delta C_T^{time\ point\ B}$). Relative fold differences were computed by $2^{-\Delta\Delta C_T}$.

RESULTS AND DISCUSSION

RKN penetrates host roots 24 hours after inoculation: The main goal of our experiment was to determine if host diurnal rhythms influence nematode invasion. Previous reports demonstrated RKN J2 penetrate *M. truncatula* after 24 hours (Dhandaydham et al, 2008). This time frame allows J2 to perceive and penetrate host root systems while minimizing developmental changes in the nematode that may alter subsequent invasions. We verified nematode RKN invasion 24 hours post inoculation in our system by selectively staining

nematodes within host root tissue using acid fuchsin (Figure 1). However, this method is non-reproducible and limits the statistical support required for robust quantification, largely due to the damage and loss of fragile root systems during staining and variations in root densities attributed to differences in cell type and age.

RKN preferentially infect host root systems during the dark: We used the sensitive and specific quantitative PCR reaction to determine RKN abundance in host root systems 24 hours after inoculation with second stage larvae. Our results indicate a diurnal bias in nematode penetration that is greatest during the dark. At 24 hours post inoculation, the nematode has perceived and migrated to a host root but has yet to feed or resume development. This allows the use of qPCR to quantitate the relative abundance of nematodes in host roots as the number of cells and nuclei per nematode will remain constant. Using a homogenous mixture of naïve J2 collected over three days after egg isolation and hatching eliminates any penetration variance attributable to 'age' of J2, post hatch. The availability of the complete genome sequence for RKN (*Meloidogyne hapla,* VW9) and its host (*Medicago truncatula*) allowed for the selection of unique, single copy, genomic loci to be selected for use in qPCR. To ensure the only variable was time of day, all other factors (temperature, moisture, etc.) were accounted and controlled; we also accounted for differences in root mass that may variably attract or otherwise facilitate nematode penetration by normalizing all nematode qPCR results (MhC217) with the qPCR results of a unique *M. truncatula* sequence (MtUBQ1). Normalization was done by subtracting the C_T values of MhC217 for a time point from the averaged C_T value of MtUBQ1 for the same time point. qPCR was conducted

on DNA isolated from whole root systems because, as expected, no discernible RKN-induced root phenotypes could be associated with nematode presence after 24 hours. Without an available DNA template, the primers did not amplify any appreciable and detectable products. Nematode abundance was sampled over four time points spanning seven hours and encompassing the change from night to day. ANOVA revealed significant differences within the data set (Table 1) and student t-test with Bonferroni multiple testing correction (Table 2) on the six possible pairwise comparisons revealed that there were consecutive significant decreases in nematode abundance after daybreak, with the highest number of nematodes in host roots to be at 2 hours before daybreak (t = -2; Figure 2). Figure 2 illustrates the normalized relative nematode abundance for the four time points; for clarity, plotted are the inverse C_T values of MhC217 normalized with MtUBQ1. The largest relative fold difference of 78.47x was seen between the time points of -2hrs and +5hrs, calculated as $2^{-\Delta\Delta C_T}$.

In order for a nematode to hatch and locate a host root, distinct signals must be perceived from a myriad of stimuli emanating from micro niches most of which are present as concentration gradients, and all of which can be expected to change drastically based on temperature and water content (Prot, 1979; Robinson, 1994). There are multiple individual root associated signals that have been implicated in both root attraction and repulsion. A single root-attraction signal is unlikely, especially when considering RKNs expansive host range. However, the comparably limited number of G-protein coupled receptors (GPCRs) in RKN (147) to *C. elegans* (1,011) (Opperman et al., 2008) suggests a restricted range of possible signals able to be perceived by RKN. The reduction of RKN GPCRs (or expansion

in *C. elegans*) certainly reflects RKN sole niche, but may also suggest that a limited number of general plant root associated signals are responsible for the root attraction phenotype. Support for a general plant associated signal is demonstrated by the lack of significant difference in RKN penetration when presented with a resistant or susceptible *Medicago* host (Dhandaydham et al., 2008).

The significant result that nematodes preferentially penetrate host root systems during the dark has is supported in current literature and emphasizes the importance of host biology to nematode parasitism. Previously, carbon dioxide (CO_2) has been described as a potent host root attractant signal (Robinson, 2004; Robinson and Perry, 2006). *Meloidogyne* spp. have also demonstrated attractiveness to moisture gradients (Prot, 1979) cumulatively these data may implicate dissolved CO_2 in the form of carbonic acid. The distinction between the gas CO_2 and carbonic acid may be critical in explaining varying results in nematode attraction assays (Bird 1962). Additionally, acidic pH was also shown to attract nematodes in media (Wang et al., 2009). Dissolved CO_2 in the form carbonic acid is a direct output of carbon metabolism and roots display diurnal growth patterns and sugar metabolism (Yanzdanbakhsh et al., 2011). Specifically, root growth and respiration is greatest during the dark (Yazdandkhsh et al., 2011), and our data implicate this diurnal cycle probably resulting in the attractive signal CO_2 (probably as carbonic acid). It is logical that nematodes have developed sensory mechanisms that detect fundamental plant signals in order to locate host root systems, rather than specific phytochemical(s) that would limit the host range of these cosmopolitan parasites. Interestingly, the majority of complex root-secreted phytochemicals

are not influenced by diurnal rhythms; although two such chemicals have been reported to demonstrate increased secretion during the transition from dark to light (Badri et al., 2010). Additionally, multiple reports describe the increase of flavonoid root exudates during the daylight (Watt and Evans, 1999; Hughes et al., 1999; Nagasaka et al., 2009). These phytochemicals may be responsible for some of the repellent activity demonstrated by some root exudates in previously reported nematode attraction assays (Diez and Dusenbery, 1988) and further suggest a ubiquitous plant signal arbitrating nematode attraction. However, other hypotheses deserve consideration, including phototaxis and maternally derived diurnal and/or circadian rhythms. In the model nematode *C. elegans*, light-sensitive photoreceptor cells are responsible for observed phototaxis, although like RKN, these nematodes supposedly exist in natural environment void of light (Ward et al., 2008). As an obligate parasite, RKN biology is certainly responsive to plant biology, and it is possible that the diurnally biased penetration of host root systems is pre-conditioned and passed maternally from one generation from the next.

The results presented here illustrate the importance of considering time and host biology when investigating pathogen-host responsiveness as many plant metabolites and signals are time dependent.

Table 1. Analysis of variance (ANOVA) of normalized qPCR results of nematode abundance

	Df	Sum Sq	Mean Sq	F value	Pr (>F)
Time	3	21.52	7.173	206.4	2.04e-12
Residuals	15	0.521	0.035		

Table 2. Student T-test of averaged normalized qPCR results of nematode abundance between time points. All pairwise comparisons are significant as determined using the conservative Bonferroni multiple testing correction ($p < 1.7E-2$).

Pairwise	**p-value**
7am - 11am	1.33E-02
7am - 12pm	2.06E-04
7am - 2pm	4.35E-04
10am - 12pm	5.18E-06
10am - 2pm	7.78E-06
12pm - 2pm	4.11E-04

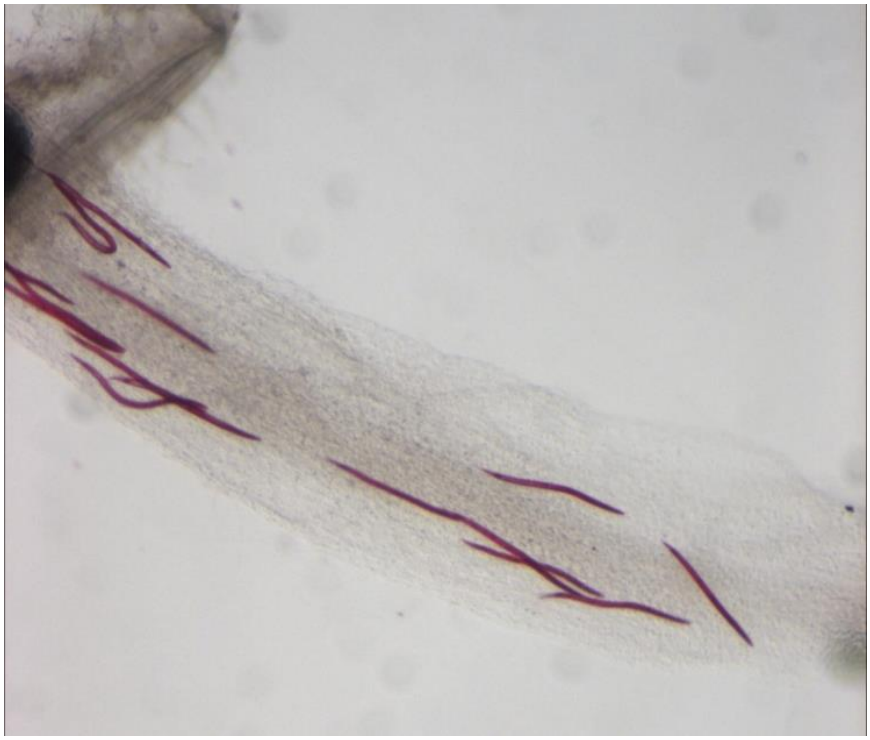


Figure 1. *Medicago truncatula* root tip infected with acid-fuchsin stained *Meloidogyne hapla* VW9 second-stage juveniles 24hours after inoculation.

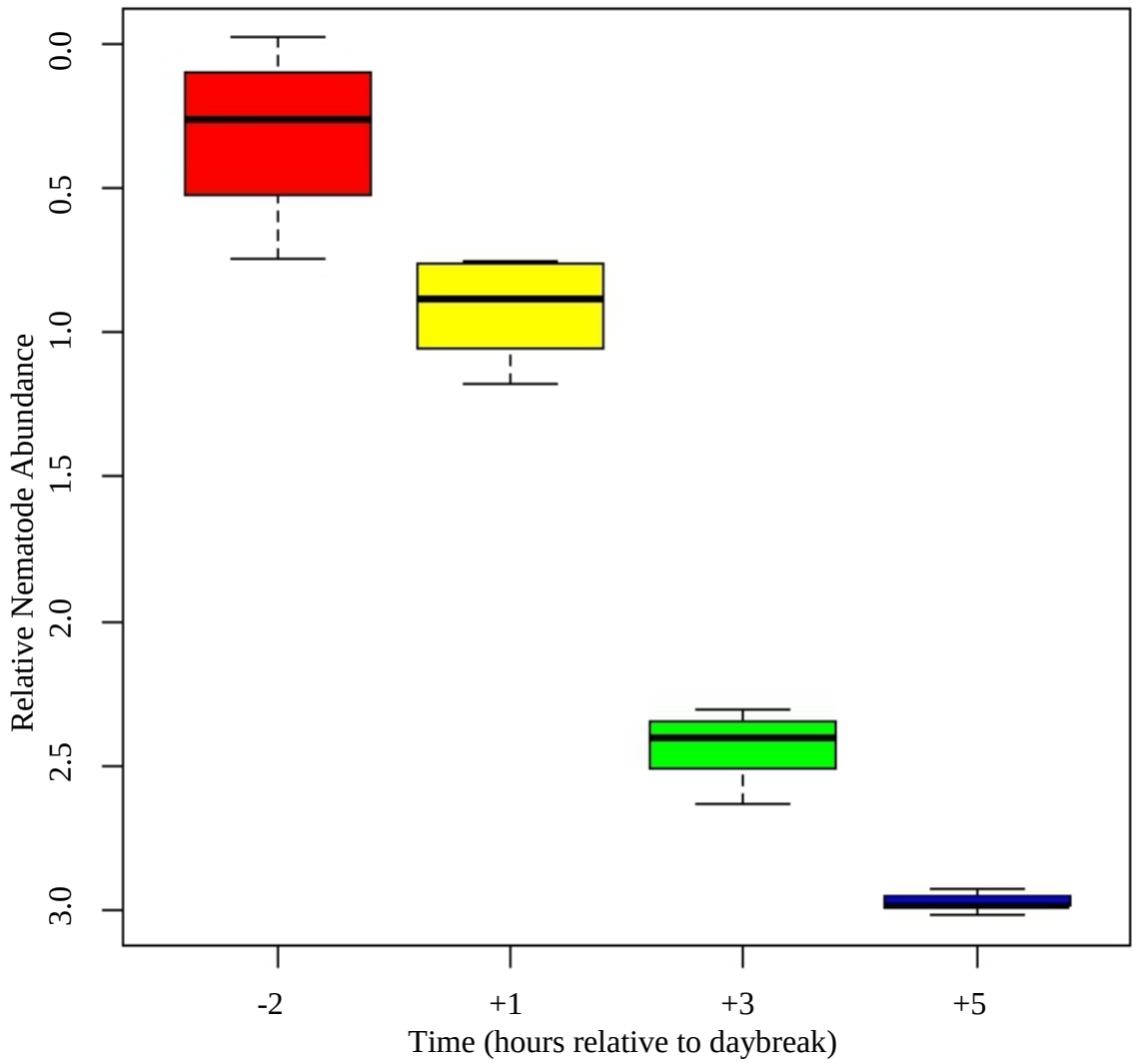


Figure 2. Box plot showing normalized relative nematode abundance as dC_T. Lower C_T values are indicative of more template DNA. Time points are relative to day break. (i.e. -2 is 2hours before day break and +1 is 1 hour after daybreak

CHAPTER 5

Postface

The broad, working hypothesis that RKN establish feeding sites by appropriating plant developmental pathways has been substantiated by the exploitation of integrated genetics, genomics, structural biology and cell biology to define the RKN-encoded signals that lie at the core of the parasitic interaction. RKN likely manipulates plant development at its most fundamental level, and this describes the ability of RKN to form their stereotypic feeding sites across vascular florae. Further, RKN likely maintain their feeding site by constantly monitoring host biology and adapting their transcriptional behavior accordingly.

At the heart of these hypotheses is the search for the signals and pathways that are central to compatible RKN-plant interactions. Although some of the components have been identified (including those presented here), understating their role in the RKN parasitic interaction remains a difficult task. The aforementioned substantial genetic system developed for RKN has greatly aided the identification of these signals, yet attempts to understand specific gene function through reverse genetics (mutagenesis, targeted gene knock-down, etc.) has yet to prove informative. For instance, RNAi knockdown of putative nematode signals involved in the host-parasite interaction has not fully blocked parasitic ability. This is in striking contrast to the situation in other plant-microbe interactions, such as the rhizobia-legume interaction, where point mutations in key receptors are sufficient to entirely block the association (Lohar and Bird, 2003). This distinction may be a consequence of RKN interacting with multiple

plant pathways or components that are either redundant, or which function in an additive manner to confer full parasitic ability. It is therefore unlikely that any one single signal would represent the linchpin to the interaction.

Despite the apparent differences between the parasitic and symbiotic symbioses, reflecting upon the constraints of host biology, there are many molecular signaling similarities between the beneficial symbiosis of legumes and rhizobia (nitrogen-fixing soil borne bacteria) with RKN infections. For instance, both symbionts induce the formation of novel structures that resemble meristems, likely by similar or identical processes. Indeed, the temporal and spatial expression patterns of two mersitematic maintenance genes (*PHAN* and *KNOX*) in the model legume *Medicago truncatula* is similar in rhizobia-induced nodules and RKN feeding sites (Koltai and Bird., 2000; Koltai et al., 2001). Many physiological similarities are also apparent between nodules and feeding sites, particularly the presence of large polyploid cells, which are consistent with the re-programming of endogenous developmental systems by these symbionts (Weerasinghe *et al.*, 2005). Further, the developmental differentiation of nodules is regulated by *ccs52*, a regulator of division arrest in mitotic cycles, producing multinucleate cells (Cebolla *et al.,* 1999) and *ENOD40* is a primary initiator of nodule formation and cortical division. The presence of *ccs52* and *ENOD40* transcripts in giant cells re-enforces the parallels between the distinct interactions (Koltai *et al*, 2001). Indeed, there are many reports of gene expression similarities between GCs and nodules (Bird and Wilson, 1994; Koltai and Bird, 2000; Koltai *et al.* 2001; Lohar *et al.* 2004. Of note to the work presented here, CLE and CEP ligands have been demonstrated to play roles in rhizobia

interactions, the formation of root galls and the allocation of host energy resources (Okamoto *et al.*, 2009; Mortier *et al.*, 2010; Imin et al., 2013), further supporting the analogous interactions of RKN and rhizobia with plants. Specifically, *ENOD40* and *ccs52* are upregulated in *CEP* overexpressing plants (unpublished data).

However, the apparent dissimilarities are not without basis, and as expected, lie at the core of both the parasitic and mutually beneficial interactions. This is best conceptualized by examining the carbon and nitrogen allocations between rhizobia and its host plant; as rhizobia fix nitrogen for use by the host, carbon and essential nutrients are provided by the host to rhizobia. Controlling the balance of this mutualistic symbiosis is a complex regulatory mechanism known as Auto-regulation Of Nodulation (AON). AON can be thought of as a communication network that regulates the requirements of the rhizobia symbiosis with the reality of the plants energy allocations. The question then remains: Does the RKN-plant interaction have a similar level of host regulation on available energy resources? As with rhizobia interactions, there is evidence that RKN are sampling photosynthates transported from the shoots. However, unlike rhizobia interactions, the impact on plant energy is more dramatic and disproportionate to the energy requirements of the nematode, even when accounting for obligate parasite biology (i.e., formation and maintenance of feeding site, and inhibition or avoidance of host immune responses, etc.). In other words, the reduction in plant yield cannot be completely accounted for by the direct requirements of the nematode (e.g. biomass of worm, egg production and gall). Another question then arises: By what mechanisms is RKN impacting plant resources? Nematodes have an estimated 60%

assimilation efficiency (Powers and McSorley, 1993), but may be as low as 20% (Melakeberhan and Ferris, 1988). On small grape plants, 3.4 kcal of energy were demanded by an *M. incognita* population over a 59-day period. This amount was equivalent to 15% of the energy assimilated by the plant (Melkakeberhan and Ferris, 1989). Yet, RKN is not a considerable competitor for plant phosphorous, removing less than 10% of total plant P (Dropkin and King, 1956). Although these data have begun to answer the questions surrounding the energy impact of RKN parasitism, the underlying principles defining this phenomenon remain unknown.

To begin investigating the mechanisms by which RKN parasitism induces drastic and disproportionate crop loss, I have initiated multiple investigations which collected comprehensive transcriptional profiles of both the host and RKN spanning time and space. The power of completed parasite and host genomes is again realized as it permits the *in silico* dissection of complex and multi-faceted tissue types, such as RKN feeding sites which contain RKN and host tissue in various degrees and stages of development. Two general approaches were taken; first, local transcriptional profiles of RKN induced knots were assayed spanning a 24 hr period and capturing host diurnal and circadian patterns. Second, global transcriptional profiles were collected from root and shoot tissue of infected and uninfected plants, spanning the initiation of the post-penetration host-parasite interaction. In short, transcriptional profiles for both RKN and the host were collected in such a way as to acquire a broad picture of host and RKN pathways responsible and responsive to a compatible interaction. These are large and complex data sets for which there are no current

algorithms suited to define them; currently, new statistical models are being developed to describe RKN pathology with these data. Here, I describe the current state of the data and analyses for these two approaches.

Circadian transcriptional oscillations in RKN: As discussed in Chapter 4, many plant processes are dependent the light-driven circadian clock, and I hypothesized that RKN biology is responsive to the rhythms and that the hosts daily energy allocations are perturbed by RKN parasitism. To examine how RKN respond to circadian rhythms, I designed an experiment to quantify the RKN and host transcriptomes over a 24 hr period. If the feeding nematode is responding transcriptionally to changes in host gene expression, this would argue that the nematode must be monitoring its host. For the pilot experiment, plants were arranged in a Complete Random Design in a growth chamber set for 18 hr days at 25°C. Tissue was collected at six time points (4 replicates per time point, 24 total samples) over a 24 hour period. This tissue was immediately frozen in liquid nitrogen and mRNA was extracted for high-throughput sequencing. Multiplexing allowed the entire experiment (24 samples) to be run on a single lane. To account for any lane bias, each of the pooled 24 samples was run on each of the 7 available lanes in the flow cell. Seven lanes of GAIIx illumina data were generated, yielding >165million, high-quality 68-mer reads. Using TopHat, 97% of these reads were mapped onto the completed genomes of *M. hapla* (6.85%) and *M. truncatula* (93.15%). Transcript abundance analysis revealed >1,000 differentially expressed loci 147 of which correspond to RKN genes. I binned these genes into 9 broad categories of expression profile (Figure 1, Appendix C). Differentially expressed plant genes

are also of interest, and using the Gene Ontology and KEGG algorithms for example, these loci can be computationally placed in the context of plant physiology.

Longitudinal and global transcriptional profiles: Also of interest are the transcriptional events that accompany the developmental changes throughout the nematode lifecycle and the subsequent host responses. To this end, *M. truncatula* was infected with *M. hapla* VW9 and tissue from uninfected and infected shoot and roots was harvested over a time course (t =1, 2, 4, 5, 7 days post inoculation) as well as *M. hapla* eggs and pre-penetration J2. RNA-seq data was collected for each of the 22 samples yielding >1billion high quality raw reads consisting of 4.37E+08 reads mapped to *M. truncatula* (86.5%) and 6.8E+07 mapped to *M. hapla* (13.5%). Currently, this large and complex data-set is being examined to better chart spatio-temporal transcriptional changes in nematode and plant. My preliminary results indicate the large majority of differentially expressed RKN genes as being 'pioneers', which exemplifies the need to develop tools to transgenically manipulate these parasites.

In the long-term the knowledge and insights gained from these works may lead to strategies to protect yield in plants via either tolerance or resistance to RKN. Plant genes encoding functions associated with resource allocation (metabolic/catabolic enzymes, transporters, channels) might be ideal targets to modulate RKN-mediated yield loss.

REFERENCES

Badri, D.V., Loyola-Vargas, V.M., Broeckling, C.D., and Vivanco, J.M. (2010). Root secretion of phytochemicals in Arabidopsis is predominantly not influenced by diurnal rhythms. Mol Plant *3*, 491–498.

Baldwin, J.G., Nadler, S.A., and Adams, B.J. (2004). Evolution of plant parasitism among nematodes. Annu Rev Phytopathol *42*, 83–105.

Bax, A. (1969). MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. J. Mag. Resonance. *65*, 355–360.

Bellafiore, S., Shen, Z., Rosso, M.-N., Abad, P., Shih, P., and Briggs, S.P. (2008). Direct Identification of the *Meloidogyne incognita* Secretome Reveals Proteins with Host Cell Reprogramming Potential. PLoS Pathog *4*, e1000192.

Bendtsen, J.D., Nielsen, H., Von Heijne, G., and Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. *340*, 783–795.

Betsuyaku, S., Takahashi, F., Kinoshita, A., Miwa, H., Shinozaki, K., Fukuda, H., and Sawa, S. (2011). Mitogen-activated protein kinase regulated by the CLAVATA receptors contributes to shoot apical meristem homeostasis. Plant Cell Physiol. *52*, 14–29.

Bird, A.F. & Loveys, B.R. (1975). The incorporation of photosynthates by Meloidogyne javanica. J. Nematol. *7*, 111-3.

Bird, A.F. 1959. The attractiveness of roots to the plant parasitic nematodes Meloidogyne javanica and M. hapla. Nematologica *4*, 322-35.

Bird, A.F., and Loveys, B.R. (1980). The involvement of cytokinins in a host–parasite relationship between the tomato (Lycopersicon esculentum) and a nematode (Meloidogyne javanica). Parasitology *80*, 497–505.

Bird, A.F., and Saurer, W. (1967). Changes Associated with Parasitism in Nematodes. II. Histochemical and Microspectrophotometric Analyses of Preparasitic and Parasitic Larvae of Meloidogyne javanica. The Journal of Parasitology *53*, 1262–1269.

Bird, D.M. (1996). Manipulation of host gene expression by root-knot nematodes. J. Parasitol *82*, 881–888.

Bird, D.M. (2004). Signaling between nematodes and plants. Current Opinion in Plant Biology *7*, 372–376.

Bird, D.M., and Wilson, M.A. (1994). DNA sequence and expression analysis of root-knot nematode-elicited giant cell transcripts. Mol. Plant Microbe Interact *7*, 419–424.

Bleckmann, A., Weidtkamp-Peters, S., Seidel, C.A.M., and Simon, R. (2010). Stem cell signaling in Arabidopsis requires CRN to localize CLV2 to the plasma membrane. Plant Physiol. *152*, 166–176.

Brand, U., Fletcher, J.C., Hobe, M., Meyerowitz, E.M., and Simon, R. (2000). Dependence of stem cell fate in Arabidopsis on a feedback loop regulated by CLV3 activity. Science *289*, 617–619.

Brunger, A.T. (2007). Version 1.2 of the Crystallography and NMR system. Nature Protocols *2*, 2728–2733.

Bybd, D.W., Kirkpatrick, T., and Barker, K.R. (1983). An Improved Technique for Clearing and Staining Plant Tissues for Detection of Nematodes. J Nematol *15*, 142–143.

Cebolla, A., Vinardell, J.M., Kiss, E., Oláh, B., Roudier, F., Kondorosi, A., and Kondorosi, E. (1999). The mitotic inhibitor ccs52 is required for endoreduplication and ploidy-dependent cell enlargement in plants. EMBO J *18*, 4476–4484.

Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr *66*, 12–21.

Chitwood, D.J. (2003). Research on plant-parasitic nematode biology conducted by the United States Department of Agriculture-Agricultural Research Service. Pest Manag. Sci *59*, 748–753.

Curtis R. H. C., Forest R., Perry R. (2009). Hatch and host location. In Root-knot nematodes (eds Perry R., Moens M., Starr J., editors.), pp. 139–162 Wallingford, UK: CAB International.

Davis, E.L., and Mitchum, M.G. (2005). Nematodes. Sophisticated Parasites of Legumes. Plant Physiol. *137*, 1182–1188.

De Boer, J.M., Yan, Y., Wang, X., Smant, G., Hussey, R.S., Davis, E.L., and Baum, T.J. (1999). Developmental expression of secretory beta-1,4-endoglucanases in the subventral esophageal glands of Heterodera glycines. Mol. Plant Microbe Interact. *12*, 663–669.

De Meutter, J., Tytgat, T., Witters, E., Gheysen, G., Van Onckelen, H., and Gheysen, G. (2003). Identification of cytokinins produced by the plant parasitic nematodes Heterodera schachtii and Meloidogyne incognita. Mol. Plant Pathol. *4*, 271–277.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR *6*, 277–293.

Dhandaydham, M., Charles, L., Zhu, H., Starr, J.L., Huguet, T., Cook, D.R., Prosperi, J.-M., and Opperman, C. (2008). Characterization of Root-Knot Nematode Resistance in Medicago truncatula. J. Nematol *40*, 46–54.

Diez, J.A., and Dusenbery, D.B. (1989). Repellent of root-knot nematodes from exudate of host roots. J Chem Ecol *15*, 2445–2455.

Ding, X., Shields, J., Allen, R., and Hussey, R.S. (1998). A secretory cellulose-binding protein cDNA cloned from the root-knot nematode (Meloidogyne incognita). MPMI. *11*, 952–959.

Dropkin, V.H., and Boone, W.R. (1966). Analysis of Host-Parasite Relationships of Root-Knot Nematodes By Single-Larva Inoculations of Excised Tomato Roots. Nematologica *12*, 225–236.

Dropkin, V.H., and King, R.C. (1956). Studies on plant parasitic nematodes homogeneously labeled with radiophosphorus. Experimental Parasitology *5*, 469–480.

Dropkin, V.H., Helgeson, J.P., and Upper, C.D. (1969). The Hypersensitivity Reaction of Tomatoes Resistant to Meloidogyne incognita: Reversal by Cytokinins. J Nematol *1*, 55–61.

Durbak, A.R., and Tax, F.E. (2011). CLAVATA signaling pathway receptors of Arabidopsis regulate cell proliferation in fruit organ formation as well as in meristems. Genetics *189*, 177–194.

Elling, A.A., Davis, E.L., Hussey, R.S., and Baum, T.J. (2007). Active uptake of cyst nematode parasitism proteins into the plant cell nucleus. Int. J. Parasitol *37*, 1269–1279.

Gao, B., Allen, R., Maier, T., Davis, E.L., Baum, T.J., and Hussey, R.S. (2001). Identification of putative parasitism genes expressed in the esophageal gland cells of the soybean cyst nematode Heterodera glycines. MPMI. *14*, 1247–1254.

Gao, B., Allen, R., Maier, T., Davis, E.L., Baum, T.J., and Hussey, R.S. (2003). The parasitome of the phytonematode Heterodera glycines. MPMI. *16*, 720–726.

Gao, B., Allen, R., Maier, T., McDermott, J.P., Davis, E.L., Baum, T.J., and Hussey, R.S. (2002). Characterisation and developmental expression of a chitinase gene in Heterodera glycines. Int. J. Parasitol. *32*, 1293–1300.

Gao, X., and Guo, Y. (2012). CLE Peptides in Plants: Proteolytic Processing, Structure-Activity Relationship, and Ligand-Receptor Interaction. J. Integrative Plant Bio. *54*, 738–745.

Glazer, I., Apelbaum, A., and Orion, D. (1985). Effect of Inhibitors and Stimulators of Ethylene Production on Gall Development in Meloidogyne javanica-Infected Tomato Roots. J Nematol *17*, 145–149.

Glazer, I., Orion, D., and Apelbaum, A. (1983). Interrelationships between Ethylene Production, Gall Formation, and Root-knot Nematode Development in Tomato Plants Infected with Meloidogyne javanica. J Nematol *15*, 539–544.

Goverse, A. and D. McK. Bird. (2011). The role of plant hormones in nematode feeding cell formation. Chapter 16 in: Genomics and Molecular Genetics of Plant-Nematode Interactions. Eds: J. Jones, G. Gheysen and C. Fenoll. Springer: Dordrecht. 557 pp.

Goverse, A., Overmars, H., Engelbertink, J., Schots, A., Bakker, J., and Helder, J. (2000). Both induction and morphogenesis of cyst nematode feeding cells are mediated by auxin. Mol. Plant Microbe Interact. *13*, 1121–1129.

Grunewald, W., Karimi, M., Wieczorek, K., Van de Cappelle, E., Wischnitzki, E., Grundler, F., Inzé, D., Beeckman, T., and Gheysen, G. (2008). A Role for AtWRKY23 in Feeding Site Establishment of Plant-Parasitic Nematodes. Plant Physiology *148*, 358 –368.

Guo, Y., Han, L., Hymes, M., Denver, R., and Clark, S.E. (2010). CLAVATA2 forms a distinct CLE-binding receptor complex regulating Arabidopsis stem cell specification. Plant J. *63*, 889–900.

Guo, Y., Ni, J., Denver, R., Wang, X., and Clark, S.E. (2011). Mechanisms of Molecular Mimicry of Plant CLE Peptide Ligands by the Parasitic Nematode Globodera rostochiensis. Plant Phys. *157*, 476 –484.

Habeck, M., Rieping, W., Linge, J.P., and Nilges, M. (2004). NOE assignment with ARIA 2.0: the nuts and bolts. Methods Mol. Biol. *278*, 379–402.

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

Hewezi, T., and Baum, T.J. (2013). Manipulation of plant cells by cyst and root-knot nematode effectors. MPMI. *26*, 9–16.

Hewezi, T., Howe, P., Maier, T.R., Hussey, R.S., Mitchum, M.G., Davis, E.L., and Baum, T.J. (2008). Cellulose binding protein from the parasitic nematode Heterodera schachtii interacts with Arabidopsis pectin methylesterase: cooperative cell wall modification during parasitism. Plant Cell *20*, 3080–3093.

Hewezi, T., Howe, P.J., Maier, T.R., Hussey, R.S., Mitchum, M.G., Davis, E.L., and Baum, T.J. (2010). Arabidopsis Spermidine Synthase Is Targeted by an Effector Protein of the Cyst Nematode Heterodera schachtii. Plant Physiol. *152*, 968–984.

Hirakawa, Y., Shinohara, H., Kondo, Y., Inoue, A., Nakanomyo, I., Ogawa, M., Sawa, S., Ohashi-Ito, K., Matsubayashi, Y., and Fukuda, H. (2008). Non-cell-autonomous control of vascular stem cell fate by a CLE peptide/receptor system. Proceedings of the National Academy of Sciences *105*, 15208–15213.

Hirsch, S., Kim, J., Muñoz, A., Heckmann, A.B., Downie, J.A., and Oldroyd, G.E.D. (2009). GRAS proteins form a DNA binding complex to induce gene expression during nodulation signaling in Medicago truncatula. Plant Cell *21*, 545–557.

Holterman, M., Holovachov, O., van den Elsen, S., van Megen, H., Bongers, T., Bakker, J., and Helder, J. (2008). Small subunit ribosomal DNA-based phylogeny of basal Chromadoria (Nematoda) suggests that transitions from marine to terrestrial habitats (and vice versa) require relatively simple adaptations. Mol. Phylogenet. Evol *48*, 758–763.

Holterman, M., Holovachov, O., van den Elsen, S., van Megen, H., Bongers, T., Bakker, J., and Helder, J. (2008). Small subunit ribosomal DNA-based phylogeny of basal Chromadoria (Nematoda) suggests that transitions from marine to terrestrial habitats (and vice versa) require relatively simple adaptations. Mol. Phylogenet. Evol *48*, 758–763.

Hothorn, M., Belkhadir, Y., Dreux, M., Dabi, T., Noel, J.P., Wilson, I.A., and Chory, J. (2011). Structural basis of steroid hormone perception by the receptor kinase BRI1. Nature *474*, 467–471.

Huang, G., Dong, R., Allen, R., Davis, E.L., Baum, T.J., and Hussey, R.S. (2006a). A root-knot nematode secretory peptide functions as a ligand for a plant transcription factor. Mol. Plant Microbe Interact *19*, 463–470.

Huang, G., Dong, R., Allen, R., Davis, E.L., Baum, T.J., and Hussey, R.S. (2006b). A root-knot nematode secretory peptide functions as a ligand for a plant transcription factor. MPMI. *19*, 463–470.

Huang, G., Gao, B., Maier, T., Allen, R., Davis, E.L., Baum, T.J., and Hussey, R.S. (2003). A profile of putative parasitism genes expressed in the esophageal gland cells of the root-knot nematode Meloidogyne incognita. Mol. Plant Microbe Interact *16*, 376–381.

Huang, G., Gao, B., Maier, T., Allen, R., Davis, E.L., Baum, T.J., and Hussey, R.S. (2003). A profile of putative parasitism genes expressed in the esophageal gland cells of the root-knot nematode Meloidogyne incognita. MPMI. *16*, 376–381.

Hughes, M., Donnelly, C., Crozier, A., and Wheeler, C.T. (1999). Effects of the exposure of roots of Alnus glutinosa to light on flavonoids and nodulation. Can. J. Bot. *77*, 1311–1315.

Imin, N., Mohd-Radzman, N.A., Ogilvie, H., Oakes, M., and Djordjevic, M.A. (2013) Nitrogen-limitation induced regulatory peptide encoding gene MtCEP1 mediates lateral root and nodule formation in Medicago truncatula. Plant Phys (*submitted*)

Jaouannet, M., Perfus-Barbeoch, L., Deleury, E., Magliano, M., Engler, G., Vieira, P., Danchin, E.G.J., Rocha, M.D., Coquillard, P., Abad, P., et al. (2012). A root-knot nematode-secreted protein is injected into giant cells and targeted to the nuclei. New Phytologist *194*, 924–931.

Johnson, B.A. (2004). Using NMRView to visualize and analyze the NMR spectra of macromolecules. Methods Mol. Biol. *278*, 313–352.

Kinoshita, A., Nakamura, Y., Sasaki, E., Kyozuka, J., Fukuda, H., and Sawa, S. (2008). Gain-of-Function Phenotypes of Chemically Synthetic CLAVATA3/ESR-Related (CLE) Peptides in Arabidopsis thaliana and Oryza sativa. Plant and Cell Physiology *49*, 999–999.

Koenning, S.R., Overstreet, C., Noling, J.W., Donald, P.A., Becker, J.O., and Fortnum, B.A. (1999). Survey of Crop Losses in Response to Phytoparasitic Nematodes in the United States for 1994. J Nematol. *31*, 587–618.

Koltai, H., and Bird, D.M. (2000). Epistatic repression of PHANTASTICA and class 1 KNOTTED genes is uncoupled in tomato. Plant J *22*, 455–459.

Koltai, H., Dhandaydham, M., Opperman, C., Thomas, J., and Bird, D. (2001). Overlapping plant signal transduction pathways induced by a parasitic nematode and a rhizobial endosymbiont. MPMI. *14*, 1168–1177.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. *305*, 567–580.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Bio. *10*, R25.

Lee, J.S., and Torii, K.U. (2012). A Tale of Two Systems: Peptide Ligand-Receptor Pairs in Plant Development. Cold Spring Harb. Symp. Quant. Biol.

Linford, M.B. (1937). Stimulated activity of natural enemies of nematodes. Science *85*, 123–124.

Linge, J.P., Habeck, M., Rieping, W., and Nilges, M. (2003). ARIA: automated NOE assignment and NMR structure calculation. Bioinformatics *19*, 315–316.

Lohar, D.P., and Bird, D.M. (2003). Lotus japonicus: a new model to study root-parasitic nematodes. Plant Cell Physiol *44*, 1176–1184.

Lohar, D.P., Schaff, J.E., Laskey, J.G., Kieber, J.J., Bilyeu, K.D., and Bird, D.M. (2004). Cytokinins play opposite roles in lateral root formation, and nematode and Rhizobial symbioses. The Plant Journal *38*, 203–214.

Mashiach, E., Schneidman-Duhovny, D., Andrusier, N., Nussinov, R., and Wolfson, H.J. (2008). FireDock: a web server for fast interaction refinement in molecular docking. Nucleic Acids Res. *36*, W229–232.

Matsubayashi, Y. (2011). Small Post-Translationally Modified Peptide Signals in Arabidopsis. Arabidopsis Book *9*.

Mbeunkui, F., Scholl, E.H., Opperman, C.H., Goshe, M.B., and Bird, D.M. (2010). Proteomic and bioinformatic analysis of the root-knot nematode Meloidogyne hapla: the basis for plant parasitism. J. Proteome Res. *9*, 5370–5381.

Melakeberhan, H., and Ferris, H. (1988). Growth and Energy Demand of Meloidogyne incognita on Susceptible and Resistant Vitis vinifera Cultivars. J Nematol *20*, 545–554.

Melakeberhan, H., and Ferris, H. (1989). Impact of Meloidogyne incognita on Physiological Efficiency of Vitis vinifera. J. Nematol. *21*, 74–80.

Meldal, B.H.M., Debenham, N.J., De Ley, P., De Ley, I.T., Vanfleteren, J.R., Vierstraete, A.R., Bert, W., Borgonie, G., Moens, T., Tyler, P.A., et al. (2007). An improved molecular phylogeny of the Nematoda with special emphasis on marine taxa. Molecular Phylogenetics and Evolution 42, 622–636.

Meng, L., and Feldman, L.J. (2010). CLE14/CLE20 peptides may interact with CLAVATA2/CORYNE receptor-like kinases to irreversibly inhibit cell division in the root meristem of Arabidopsis. Planta *232*, 1061–1074.

Meng, L., Ruth, K.C., Fletcher, J.C., and Feldman, L. (2010). The Roles of Different CLE Domains in Arabidopsis CLE Polypeptide Activity and Functional Specificity. Mol. Plant *3*, 760–772.

Mitchum, M.G., Wang, X., and Davis, E.L. (2008). Diverse and conserved roles of CLE peptides. Curr. Op. in Plant Bio. *11*, 75–81.

Mortier, V., Den Herder, G., Whitford, R., Van de Velde, W., Rombauts, S., D'haeseleer, K., Holsters, M., and Goormachtig, S. (2010). CLE Peptides Control Medicago truncatula Nodulation Locally and Systemically. Plant Physiol. *153*, 222–237.

Müller, R., Bleckmann, A., and Simon, R. (2008). The Receptor Kinase CORYNE of Arabidopsis Transmits the Stem Cell–Limiting Signal CLAVATA3 Independently of CLAVATA1. Plant Cell *20*, 934–946.

Nagasaka, S., Takahashi, M., Nakanishi-Itai, R., Bashir, K., Nakanishi, H., Mori, S., and Nishizawa, N.K. (2009). Time course analysis of gene expression over 24 hours in Fe-deficient barley roots. Plant Mol Biol *69*, 621–631.

Ni, J., Guo, Y., Jin, H., Hartsell, J., and Clark, S.E. (2011). Characterization of a CLE processing activity. Plant Mol. Biol. *75*, 67–75.

Nielsen, H., and Krogh, A. (1998). Prediction of signal peptides and signal anchors by a hidden Markov model. Proc Int Conf Intell Syst Mol Biol *6*, 122–130.

Nielsen, H., Engelbrecht, J., Brunak, S., and Von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. *10*, 1–6.

Ohyama, K., Ogawa, M., and Matsubayashi, Y. (2008). Identification of a biologically active, small, secreted peptide in Arabidopsis by in silico gene screening, followed by LC-MS-based structure analysis. Plant J. *55*, 152–160.

Okamoto, S., Ohnishi, E., Sato, S., Takahashi, H., Nakazono, M., Tabata, S., and Kawaguchi, M. (2009). Nod Factor/Nitrate-Induced CLE Genes that Drive HAR1-Mediated Systemic Regulation of Nodulation. Plant Cell Physiol. *50*, 67–77.

Opperman, C.H., Bird, D.M., Williamson, V.M., Rokhsar, D.S., Burke, M., Cohn, J., Cromer, J., Diener, S., Gajan, J., Graham, S., et al. (2008). Sequence and genetic map of

Meloidogyne hapla: A compact nematode genome for plant parasitism. PNAS. *105*, 14802 – 14807.

Phillips, M., Jones, J., and Furlanetto, C. (2007). The role of flavonoids produced in response to cyst nematode infection of Arabidopsis thaliana. Nematology *9*, 671–677.

Powers, L.E., and McSorley, R. (1993). Energetics of Meloidogyne incognita on Resistant and Susceptible Alyceclover Genotypes. J Nematol *25*, 257–264.

Prot, J.C. (1980). Migration of plant-parasitic nematodes towards plant roots. Revue de Nematologie *3*, 305–318.

Qin, L., Kudla, U., Roze, E.H.A., Goverse, A., Popeijus, H., Nieuwland, J., Overmars, H., Jones, J.T., Schots, A., Smant, G., et al. (2004). Plant degradation: a nematode expansin acting on plants. Nature *427*, 30.

Replogle, A., Wang, J., Paolillo, V., Smeda, J., Kinoshita, A., Durbak, A., Tax, F.E., Wang, X., Sawa, S., and Mitchum, M.G. (2013). Synergistic interaction of CLAVATA1, CLAVATA2, and RECEPTOR-LIKE PROTEIN KINASE 2 in cyst nematode parasitism of Arabidopsis. MPMI. *26*, 87–96.

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. *16*, 276–277.

Riddle, D.L., and Bird, A.F. (1985). Responses of the plant parasitic nematodes Rotylenchulus reniformis, Anguina agrostis and Meloidogyne javanica to chemical attractants. Parasitol. *91 (Pt 1)*, 185–195.

Robinson AF, Perry RN. (2006). Behavior and Sensory Perception. In: Perry RN, Moens M, editors. Plant Nematology. Wallingford, UK: CABI. pp. 210–233.

Robinson, A.F. (1994). Movement of Five Nematode Species through Sand Subjected to Natural Temperature Gradient Fluctuations. J. Nematol. *26*, 46–58.

Rosso, M.N., Favery, B., Piotte, C., Arthaud, L., De Boer, J.M., Hussey, R.S., Bakker, J., Baum, T.J., and Abad, P. (1999). Isolation of a cDNA encoding a beta-1,4-endoglucanase in the root-knot nematode Meloidogyne incognita and expression analysis during plant parasitism. MPMI. *12*, 585–591.

Sasser, J.N. (1980). Root-knot nematodes: a global menace to crop production. Plant Disease *64*, 36–41.

Sasser, J.N., and Carter, C.C. (1985). An Advanced Treatise on Meloidogyne: Biology and Control (North Carolina State University, Department of Plant Pathology).

Schaff, J.E., Mbeunkui, F., Blackburn, K., Bird, D.M., and Goshe, M.B. (2008). SILIP: a novel stable isotope labeling method for in planta quantitative proteomic analysis. Plant J. *56*, 840–854.

Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., and Wolfson, H.J. (2005). PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic Acids Res. *33*, W363–367.

Sharma, V.K., Ramirez, J., and Fletcher, J.C. (2003). The Arabidopsis CLV3-like (CLE) genes are expressed in diverse tissues and encode secreted proteins. Plant Mol. Biol *51*, 415–425.

Smant, G., Stokkermans, J.P.W.G., Yan, Y., De Boer, J.M., Baum, T.J., Wang, X., Hussey, R.S., Gommers, F.J., Henrissat, B., Davis, E.L., et al. (1998). Endogenous cellulases in animals: Isolation of β-1,4-endoglucanase genes from two species of plant-parasitic cyst nematodes. PNAS. *95*, 4906 –4911.

Smith, T.F., and Waterman, M.S. (1981). Identification of common molecular subsequences. J. Mol. Biol. *147*, 195–197.

Stone, E.F., Fulton, B.O., Ayres, J.S., Pham, L.N., Ziauddin, J., and Shirasu-Hiza, M.M. (2012). The Circadian Clock Protein Timeless Regulates Phagocytosis of Bacteria in Drosophila. PLoS Pathog *8*, e1002445.

Strabala, T.J. (2008). CLE genes in plant development. Plant Signal Behav *3*, 457–459.
Teillet, A., Dybal, K., Kerry, B.R., Miller, A.J., Curtis, R.H.C., and Hedden, P. (2013). Transcriptional Changes of the Root-Knot Nematode Meloidogyne incognita in Response to Arabidopsis thaliana Root Signals. PLoS ONE *8*, e61259.

Ten Hove, C.A., Bochdanovits, Z., Jansweijer, V.M.A., Koning, F.G., Berke, L., Sanchez-Perez, G.F., Scheres, B., and Heidstra, R. (2011). Probing the roles of LRR RLK genes in Arabidopsis thaliana roots using a custom T-DNA insertion set. Plant Mol. Biol. *76*, 69–83.

Thomas, V.P., Fudali, S.L., Schaff, J.E., Liu, Q., Scholl, E.H., Opperman, C.H., Bird, D.M., and Williamson, V.M. (2012). A sequence-anchored linkage map of the plant-parasitic nematode Meloidogyne hapla reveals exceptionally high genome-wide recombination. G3 (Bethesda) *2*, 815–824.

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

Viglierchio, D.R., and Lownsbery, B.F. (1960). The Hatching Response of Meloidogyne Species To the Emanations From the Roots of Germinating Tomatoes. Nematologica *5*, 153–157.

Wang, C., Bruening, G., and Williamson, V.M. (2009). Determination of Preferred pH for Root-knot Nematode Aggregation Using Pluronic F-127 Gel. J Chem Ecol *35*, 1242–1251.

Wang, X., Allen, R., Ding, X., Goellner, M., Maier, T., De Boer, J.M., Baum, T.J., Hussey, R.S., and Davis, E.L. (2001). Signal peptide-selection of cDNA cloned directly from the esophageal gland cells of the soybean cyst nematode Heterodera glycines. MPMI. *14*, 536–544.

Wang, X., Mitchum, M.G., Gao, B., Li, C., Diab, H., Baum, T.J., Hussey, R.S., And Davis, E.L. (2005). A parasitism gene from a plant-parasitic nematode with function similar to *CLAVATA3/ESR (CLE)* of Arabidopsis thaliana. Mol. Plant Path. *6*, 187–191.

Ward, A., Liu, J., Feng, Z., and Shawn Xu, X.Z. (2008). Light-sensitive neurons and channels mediate phototaxis in C. elegans. Nat Neurosci *11*, 916–922.

Watt, M., and Evans, J.R. (1999). Linking Development and Determinacy with Organic Acid Efflux from Proteoid Roots of White Lupin Grown with Low Phosphorus and Ambient or Elevated Atmospheric CO2 Concentration. Plant Physiol. *120*, 705–716.

Weerasinghe, R.R., Bird, D.M., and Allen, N.S. (2005). Root-knot nematodes and bacterial Nod factors elicit common signal transduction events in Lotus japonicus. Proceedings of the National Academy of Sciences of the United States of America *102*, 3147 –3152.

Weerasinghe, R.R., Bird, D.M., and Allen, N.S. (2005). Root-knot nematodes and bacterial Nod factors elicit common signal transduction events in Lotus japonicus. PNAS. *102*, 3147 – 3152.

Whitford, R., Fernandez, A., De Groodt, R., Ortega, E., and Hilson, P. (2008). Plant CLE peptides from two distinct functional classes synergistically induce division of vascular cells. Proceedings of the National Academy of Sciences *105*, 18625–18630.

Wubben, M.J., 2nd, Su, H., Rodermel, S.R., and Baum, T.J. (2001). Susceptibility to the sugar beet cyst nematode is modulated by ethylene signal transduction in Arabidopsis thaliana. Mol. Plant Microbe Interact. *14*, 1206–1212.

Wüthrich, K., NMR of proteins and nucleic acids. The George Fisher Baker non-resident lectureship in chemistry at Cornell University. 1986, New York: Wiley. xv, 292.

Wyss, U., and Grundler, F.M.W. (1992). Feeding behavior of sedentary plant parasitic nematodes. Netherlands J. of Plant Path. *98*, 165–173.

Wyss, U., and U. Zunke. (1986). Observations on the behaviour of second-stage juveniles of *Heterodera schachtii* inside host roots. Revue de Nématologie *9*,153–165.

Xue, B., Hamamouch, N., Li, C., Huang, G., Hussey, R.S., Baum, T.J., and Davis, E.L. (2013). The 8D05 parasitism gene of Meloidogyne incognita is required for successful infection of host roots. Phytopathology *103*, 175–181.

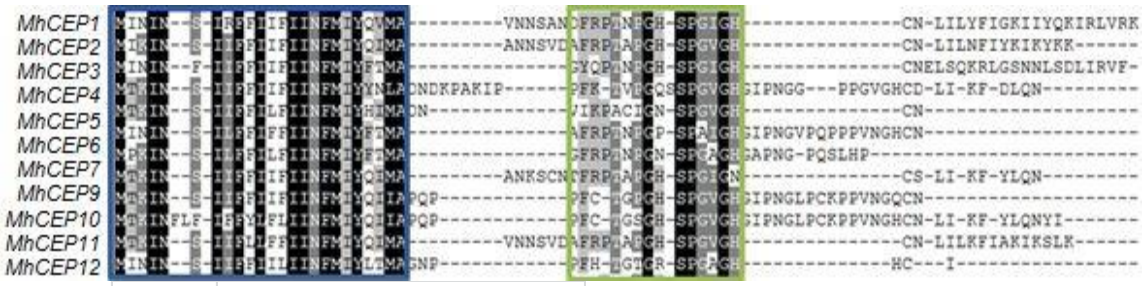
Yazdanbakhsh, N., Sulpice, R., Graf, A., Stitt, M., and Fisahn, J. (2011). Circadian control of root elongation and C partitioning in Arabidopsis thaliana. Plant Cell Environ. *34*, 877–894.

APPENDICES

APPENDIX A

Chapter 2 Supplemental Information

(a)



(b)

Gene	ID Genomic Location	
MhCEP1	MhA1_Contig363:16662..16859	-DFRPTNPGHSPGIGH
MhCEP2	MhA1_Contig363:19270..19450	-AFRPTAPGHSPGVGH
MhCEP3	MhA1_Contig363:21115..21294	-GYQPTNPGHSPGIGH
MhCEP4	MhA1_Contig363:22230..22439	-PFKTVPGQSSPGVGH
MhCEP5	MhA1_Contig363:23916..24047	-VIKPACIGNSPGVGH
MhCEP6	MhA1_Contig363:25332..25502	-AFRPTNPGPSPAIGH
MhCEP7	MhA1_Contig363:27659..27811	-GFRPTNPGNSPGAGH
MhCEP8	MhA1_Contig363:29266..29430	-TFRPTAPGHSPGIGN
MhCEP9	MhA1_Contig363:31159..31332	-PFCTGPGH-SPGVGH
MhCEP10	MhA1_Contig368:212..1	-PFCTGSGH-SPGVGH
MhCEP11	MhA1_Contig368:3206..3028	-AFRPTAPGHSPGVGH
MhCEP12	MhA1_Contig368:6989-6858	-PFHTGTGR-SPGAGH
MtCEP1	Mtr.7265.1.S1_at, MTEST_TC169364	-AFQPTTPGNSPGVGH
MtCEP2	MTEST_META519TF	-AFRPTTPGHSPGIGH
MtCEP3	MTEST_TC125059	-DFRPTTPGNSPGAGH
MtCEP4	Medtr5g025790.1	-AFRPTSPGHSPGVGH
MtCEP5	Medtr5g017710.1	-AFRPTTPGHSPGVGH
MtCEP6	AC233112_1004.1	-AFRPTSPGHSPGVGH
MtCEP7	AC233112_1013.1	-SFRPTTPGSSPGVGH
MtCEP8	AC233112_1014.1	-AFRPTTPGNSPGVGH
MtCEP9	AC233112_1015.1	-AFRPTTPGSSPGVGH
MtCEP10	Medtr5g030490.1	-AFRPTPSGHSIGVGH
MtCEP11	Medtr8g086660.1	-AFRPTSPGHSPGVGH
AtCEP1	At1g47485	-DFRPTNPGNSPGVGH
AtCEP2	At1g59835	-DFAPTNPGDSPGIRH
AtCEP3	At2g23440	DTFRPTEPGHSPGIGH
AtCEP4	At2g35612	DAFRPTHQGPSQGIGH
AtCEP5	At5g66815	-DFRPTTPGHSPGIGH

Figure S1. Complete RKN CEP gene sequences and plant CEP ligand domains

(a) Full *M. hapla* CEP mimic sequences; secretions signal domain (blue box) nearly adjacent to active domain (green box). (b) An alignment of *M. hapla* (MhCEP), *M. truncatula* (MtCEP) and *A. thaliana* (AtCEP) active domains reveals high sequence similarity, and biochemical properties of individual residues.

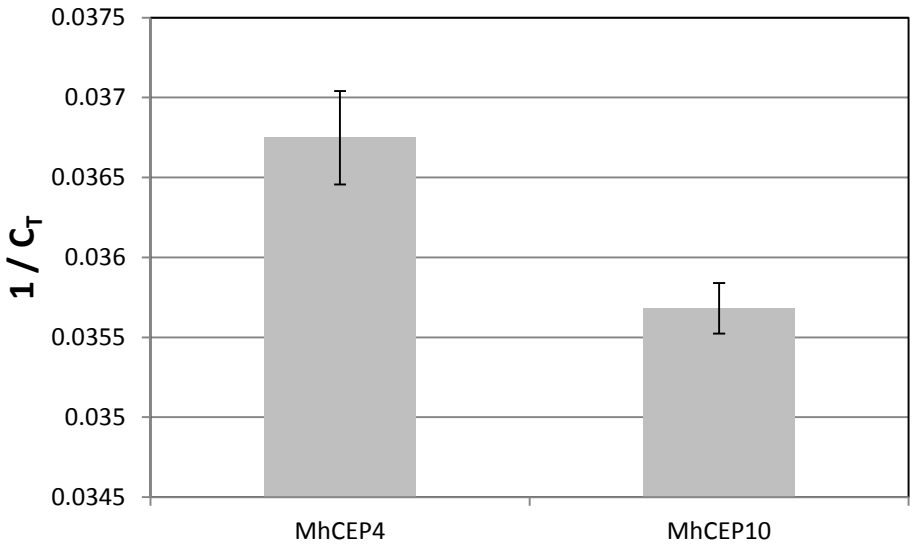
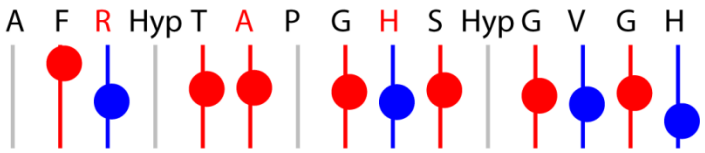


Figure S2. qRT-PCR validated RNA-Seq expression of *MhCEP4* and *MhCEP10***.** Unique and specific primers could only be designed for *MhCEP4* and *MhCEP10.* Plotted are the inverse average C_T of 3 replicates, representing relative gene expression, with standard deviation error bars. RNA was isolated from three-week post-inoculation RKN induced knots. These results confirm the variable expression pattern observed between *MhCEP4* and *MhCEP10* in the RNA-Seq data.

(a)



(b)

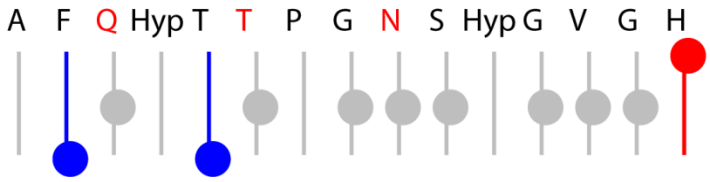


Figure S3. Chemical Shift Indices for MhCEP11 and MtCEP1**.**

The chemical shift indicies are shown for MhCEP11 (A) and MtCEP1 (B) and indicate the propensity for secondary structure within a sequence. Chemical shift indicies are based on the chemical shifts of the HN atoms compared to a database of structures. Red dots denoted alpha helical character, blue dots denote b-strand character while grey dots denote random coil. The location of the dot on the vertical scale denotes the level of certainty. Clearly from this figure, MhCEP11 (panel A) has a greater propensity to form alpha helical character compared to that of MtCEP1.

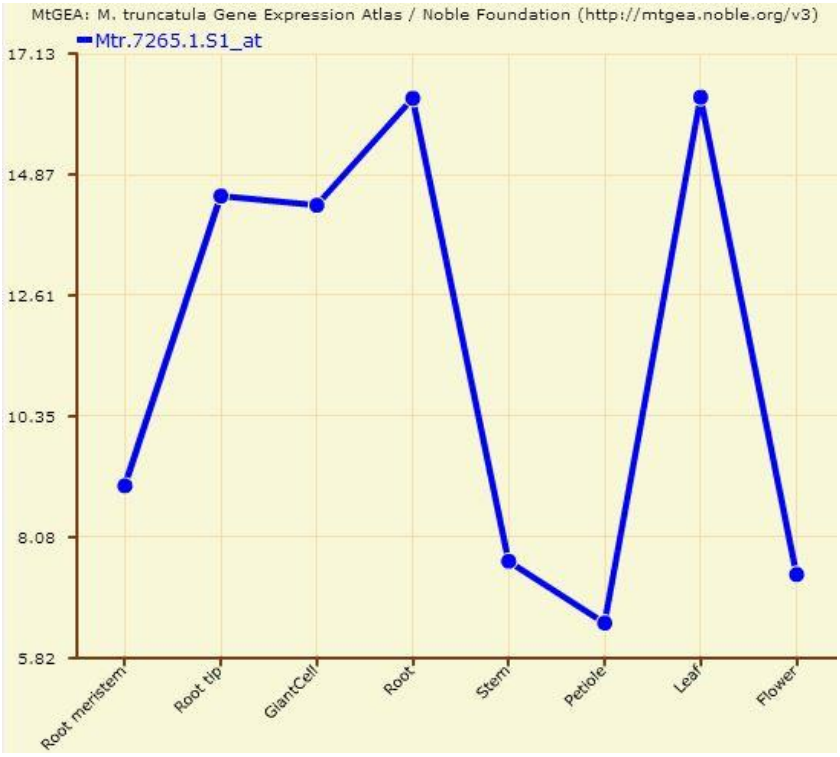


Figure S4. *MtCEP1* is expressed in diverse tissues**.**

Examination of *MtCEP1* (accession Mtr.7265.1.S1_at) expression using MtGEA (mtgea.noble.org/) shows relatively high expression in roots and leaf tissues.

APPENDIX B

Chapter 3 Supplemental Information

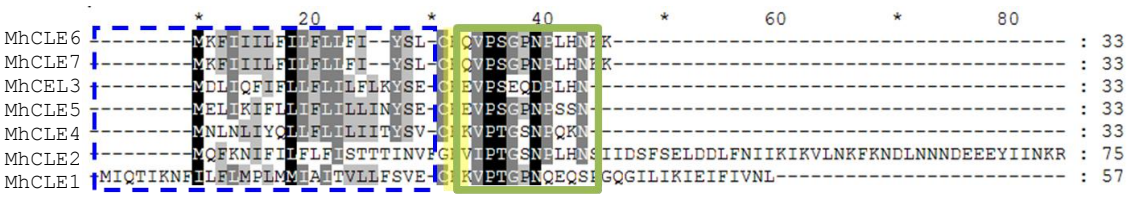


Figure S1. Complete protein sequence of the seven *M. hapla* CLE mimics.

Blue box; conserved nematode secretion signal sequence. Yellow; predicted signal sequence cleavage site. Green box; conserved CLE ligand domain.

APPENDIX C

Chapter 5 Preliminary Data

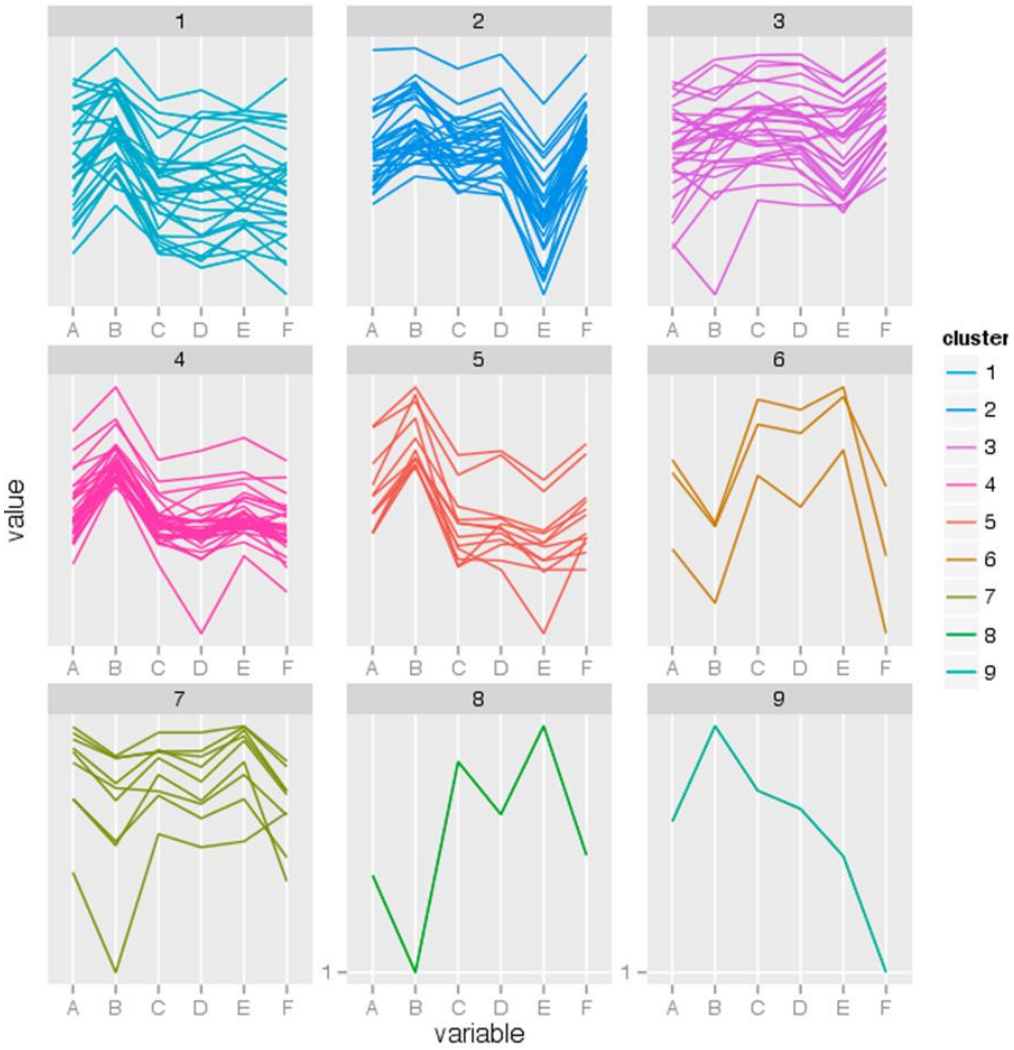


Figure 1. Expression profiles of *M. hapla* genes genes that are differentially expressed over a 24 hour period. Differentially expressed genes were binned into nine different expression patterns. Time points A, B, C, D, E, and F represent 10:30 pm, 2:00 am, 5:00 am, 6:30 am, 2:00 pm and 9:00 pm respectively.