De novo analysis of the transcriptome of *Pratylenchus zeae* to identify transcripts for proteins required for structural integrity, sensation, locomotion and parasitism

John Fosu-Nyarko[1, 2*], Jo-Anne C. H. Tan[1], Reetinder Gill[1], Vaughan G. Agrez[1], Uma Rao[3], Michael G. K. Jones[1*]

[1]*Plant Biotechnology Research Group, WA State Agricultural Biotechnology Centre, School of Veterinary and Life Sciences, Murdoch University, Perth, WA 6150, Australia,*

[2]*Nemgenix Pty Ltd, WA State Agricultural Biotechnology Centre, Murdoch University, Perth, WA 6150, Australia*

[3]*Division of Nematology, Indian Agricultural Research Institute, New Delhi, 110012, India*

***Corresponding authors:**

John Fosu-Nyarko

Email: J.Fosu-Nyarko@murdoch.edu.au; Phone: +61 8 93602920; fax: +61 8 93602502

Michael G.K. Jones

Email: M.Jones@murdoch.edu.au; Phone: +61 8 93602424; fax: +61 8 93602502

Running title: *Pratylenchus zeae* transcriptome

Keywords: Transcriptome, *Pratylenchus* spp, parasitism genes, root lesion nematode, nematode effectors.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/mpp.12301

Accession numbers: BioProject ID PRJNA268047; SRR1657910 (Short Read Archive); SUB990140 (Transcriptome Shotgun Assembly temporary BioSample submission ID).

Summary

The root lesion nematode *Pratylenchus zeae,* a migratory endoparasite, is an economically important pest of major crop plants (e.g. cereals, sugarcane). It enters host roots, migrates through root tissues and feeds from cortical cells, and defends itself against biotic and abiotic stresses in the soil and in host tissues. We report *de novo* sequencing of the *P. zeae* transcriptome using 454 FLX, and identification of putative transcripts encoding proteins required for movement, response to stimuli, feeding and parasitism. Sequencing generated 347,443 good quality reads which were assembled into 10,163 contigs and 139,104 singletons: 65% of contigs and 28% of singletons matched sequences of free-living and parasitic nematodes. Three quarters of annotated transcripts were common to reference nematodes, mainly representing genes encoding proteins for structural integrity and fundamental biochemical processes. Over 15,000 transcripts were similar to *Caenorhabditis elegans* genes encoding proteins with roles in mechanical and neural control of movement, responses to chemicals, mechanical and thermal stresses. Notably, 766 transcripts matched parasitism genes employed by both migratory and sedentary endoparasites in host interactions, of which three studied hybridised to the gland cell region, suggesting that they might be secreted. Conversely, transcripts for effectors reported to be involved in feeding site formation by sedentary endoparasites were conspicuously absent. Transcripts similar to those encoding some

secretory-excretory products at the host interface of *Brugia malayi*, the secretome of *Meloidogyne incognita* and products of gland cells of *Heterodera glycines* were also identified. This *P. zeae* transcriptome provides new information for genome annotation and functional analysis of possible targets for control of pratylenchid nematodes.

Introduction

Pratylenchus zeae is a migratory endoparasitic root lesion nematode (RLN) present in many agricultural regions. It can infect major economically important crops such as cereals, fruits and vegetables, cotton, coffee and sugarcane, so contributing to substantial yield losses (Castillo and Volvas 2007; Blair and Stirling 2007). *P. zeae* infestation is characterised by plant stunting, wilting, premature leaf yellowing, poor root development and the presence of brown lesions on roots. These symptoms result from root damage causing nutritional and water stress, and from physical damage caused by nematode migration and feeding. Root damage also enables other soil microorganisms and root pathogens to enter (Khan 1959). Control of plant parasitic nematodes (PPNs) normally involves resistance breeding, cultural practices or application of chemical nematicides: for broadscale agriculture chemical control is usually uneconomical, and many of the older nematicides have now been banned or their use restricted because of environmental and human risk concerns (Chitwood 2002). This has prompted a search for alternative methods of control, including biological control agents, more environmentally benign chemicals, or new forms of genetic control. Delivery of compounds that prevent growth of PPNs via transgenic

plants is one such approach, e.g. by expressing proteins that inhibit nematode behavioural or digestive functions, or RNA interference (RNAi) to inactivate genes vital for nematode parasitism or metabolism. The identification of suitable target genes for such control strategies requires detailed knowledge of the genes present and their function, which until recently has been lacking (Jones and Fosu-Nyarko 2014; Fosu-Nyarko and Jones 2015).

The life cycle of *Pratylenchus* species varies from 3-8 weeks depending on conditions and host (Castillo and Vovlas 2007). *P. zeae* reproduces by parthenogenesis: adult females lay eggs singly or in small groups, first stage juveniles (J1) develop in the eggs and moult to the first infective J2 stage, followed by three further moults (to J3, J4 and adults). All stages from J2s to adults are vermiform and motile, can leave and enter roots, and can feed from host cells (Stirling 1991; Jones and Fosu-Nyarko 2014). They locate host roots via gradients in the soil rhizosphere (Trevathan et al. 1985), then enter roots, move from cell-to-cell and feed from cell cytoplasm using their mouth stylet. Feeding is accompanied with mechanical probing and secretion of compounds and effectors from gland cells, chemosensory sensilla, and amphids (Zunke 1990; Perry 1996; Reynolds et al. 2011).

Partly because RLNs do not induce permanent feeding sites such as giant cells, syncytia or nurse cells (Jones 1981), which enable infection sites of sedentary endoparasites to be identified readily, host-pathogen studies of RLNs have been neglected despite their economic importance. With new genomic technologies this situation is now changing, and it is evident that RLNs are amenable to RNAi: down-

regulation of genes required for movement significantly decrease their survival and reproduction (Soumi et al. 2012; Tan et al. 2013). RNAi studies in PPNs have two complementary roles: to study gene function and to define targets for an RNAi strategy to control them. Since RLNs actively seek and feed from host cells throughout their lives, genes required for locomotion, neuro-reception (e.g. chemo- and thermo-reception) and successful parasitism are expressed in all infective stages, making them good subjects for functional studies. However, to date there has been no work to identify and analyse such transcripts/genes for *P. zeae*. Here, we report sequencing and analysis of the transcriptome of mixed stages of *P. zeae,* to identify transcripts for proteins that make them successful parasites. These include transcripts for proteins similar to those required for movement and response to stimuli in other nematodes, and putative transcripts encoding effectors similar to those secreted from gland cells by other PPNs. Such effectors may modify cell walls, aid ingestion and digestion of host cell cytoplasm, and enable them to evade or neutralise host immune responses. Identification and functional analysis of such transcripts is needed to understand why RLNs are widespread and successful crop pests, and how their genome complement and secreted effectors function and differ from those of sedentary endoparasites. This knowledge will contribute to developing new strategies for genetic or chemical control of these important agricultural pests (Fosu-Nyarko and Jones 2015).

Results

P. zeae transcriptome

Sequencing of the transcriptome of motile J2 to adult stages of *P. zeae* yielded 347,443 high quality reads: over 98% had an average PHRED score of 22 or more. The reads consisted of 32,602,958 nucleotides (nt) with a mean read length of 178.8 nt, 60% (208,429) of which were assembled using the SoftGenetics NextGEne V2.16 into 10,163 contigs with an average of 17.9 reads/contig. The minimum number of reads/contig was 2 and the maximum was 5,746 (for contig251, 450 nt long), 11 other contigs were composed of a similar number of reads. The contigs ranged in size from 16 to 2,221 nt with an average of 470.5 nt and an N50 of 519 bases. About 72% of the contigs were 200 – 599nt, with the highest number between 300 – 399nt (Fig. 1). The singletons (139,015 of the reads not used in contig assembly) ranged from 40 to 752 nt, with an average length of 182 nt. The A+T content of contigs and singletons were 53.9% and 55.3% respectively. Transcripts <100 nt long (53 contigs and 46,044 singletons) were excluded from functional analyses. The reads and assembled transcripts have been deposited at the National Center for Biotechnology Information (NCBI) under BioProject ID PRJNA268047; with the accession number SRR1657910 for deposits in the Short Read Archive and temporary BioSample submission ID; SUB990140 for the assembled transcripts in the Transcriptome Shotgun Assembly database.

Analysis of transcripts putatively encoding CAZymes and identification of possible contamination

Considerable care was taken to prevent possible contamination of the starting material and during the sequencing process, since this can be an issue for high throughput sequencing data (Lusk 2014). Possible contamination in the *P. zeae* transcriptome was investigated using Alien Index (AI) analysis as described by Gladyshev et al. (2008) and Eves-van den Akker et al. (2014). To do this, the AI was calculated for all transcripts (singletons and contigs) that returned a BLASTX hit to at least a sequence of a metazoan or a non-metazoan species in the NCBI non-redundant nr/nt database at an e-value threshold of 1E-05. An AI was not calculated for transcripts without a hit to a sequence in the database. Transcripts with an AI >0 indicated a better hit to a non-metazoan species than to a metazoan species: these were considered as possible contaminants and on this basis 796 of these, including 64 contigs, were excluded from further analysis. This analysis was particularly important before comparison of the transcripts to Carbohydrate active enzymes (CAZymes) in the CAZy database (www.cazy.org), which also contain CAZymes of bacterial and fungal origin.

The recent compilation of sequences of CAZymes (CAZy database, November, 2014) which contained 188, 123 protein sequences of families of glycoside hydrolases (GHs) including carbohydrate binding modules (CBMs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), and those for Auxiliary Activities (AA) were downloaded and together with sequences of 500 glycosyltransferases (GTs) were compared to the *P. zeae* transcripts using a local BLASTX on CLC Genomics Workbench 7.5. At thresholds of

1E-5 *e*-value and BLAST hit per cent identity of 50, 646 transcripts were identical to known CAZymes. For any transcript with similarity to those of CAZymes of both metazoans and non-metazoans, a further filter was used to assess its possible 'foreignness' to the transcriptome. For this, if a transcript had ≥50% BLASTX hit identity to a non-metazoan sequence, but less than 40% identity to any nematode sequence, it was excluded from the analysis. This resulted in the identification of 607 transcripts with matches to sequences of the following families of CAZymes: 20 GHs, one PL3, four CBMs, two CEs, 33 GTs and no AAs (Table S1). A majority of the transcripts matched CAZymes of both parasitic and free-living nematodes, mainly *C. elegans*, *Strongloides ratti,* and *Ascaris suum*, and were identified as involved in common molecular processes of development e.g. endoplasmic reticulum degradation-enhancing alpha-mannosidase, trehalases and lysozymes (Table S1).

Sixty three *P. zeae* transcripts matched CAZymes of plant parasitic nematodes (Table 1). Some transcripts matched CAZymes with multiple enzyme domains of the different classes, and these were generally GHs with CBMs. For example, 27 transcripts matched GH5 cellulases of *Ditylenchus destructor, Pratylenchus coffeae* and *Radopholus similis,* and these proteins are also known to have CBM2 activity (Table 1). A total of 45 transcripts matched GH5 cellulases of five genera of PPNs; these were *Pratylenchus* species (*P. penetrans, P. pratensis, P. vulnus* and *P. coffeae*), the lesion nematode *R. similis, Aphelenchoides fragariae, Rotylenchulus reniformis* and *D. destructor*. Two transcripts were also similar to those of PL3 pectate lyases of the cyst nematodes *H. schachtii* and *H. glycines*. Together with GH5 cellulases, pectate lyases are well-

characterised cell wall modifying enzymes secreted from gland cells of PPNs during the migration phase of host invasion (Smant et al. 1998).

A total of 52 transcripts best matched the GH family of CAZymes of seven non-nematode origins; *Oryza sativa* and *Oryza indica*, the bacteria *Peptoclostridium difficile* and fungi *Piriformospora indica* and *Leptosphaeria maculans*, and the insect *Nilaparvata lugens.* Further analyses of these transcripts indicate they also had ≥ 40% identity to similar nematode proteins. For example, the GH18 protein LEMA_P082410.1 of *Leptosphaeria maculans* JN3, whose transcripts are putatively similar to 32 of *P. zeae*, also has 55% sequence identity to protein kinases of both free-living and parasitic nematodes including *Ceanorhabditis japonica, Pristionchus pacificus, Brugia malayi* and *Heamonchus contortus*.

The majority (75%) of the 344 transcripts matching the 34 glycosyl transferase families were mainly to those of the nematodes *C. elegans* and *A. suum,* and are required for common biological or developmental processes (Table S1). The best matches to plant parasitic nematodes included GT2 chitin synthase of the root knot nematode *Meloidogyne artiellia* and GT20 putative trehalose 6-phosphate synthase of *Aphelenchus avenae*. In addition, there were matches to exostosin-1 and 2 (of *A. suum*), an endoplasmic reticulum-resident type II transmembrane glycosyltransferase involved in biosynthesis of heparan sulfate. Interestingly, there were transcripts similar to those encoding *bre-3* and *bre-4* of *C. elegans* and *bre-4* and *bre-5* of *A. suum,* which encode GT2 beta-1,4-mannosyltransferases, and in the case of *C. elegans,* are expressed in the gut epithelium and are required for resistance to toxicity of *Bacillus thuringiensis* Cry5B

(Marroquin et al. 2000). For 69 transcripts the best matching GT CAZymes were of non-nematode origin, however, they also had ≥40% BLASTX hit identity to several similar nematodes proteins.

In silico functional annotation of transcripts

About 48% of contigs and 15% of singletons were assigned putative functions from their similarities to genes encoding proteins involved in biological, molecular and/or cellular processes using the Gene Ontology (GO) classification and genetic information from *Caenorhabditis elegans*. These (4,827 contigs and 21,129 singletons) matched genes associated with 912 GO terms, including 420 multiple terms associated with biological functions, 321 terms with molecular functions and 171 terms with cellular functions (Fig. 2). A total of 52% of the transcripts had putative molecular functions associated with ligand binding – a term which describes how molecules including secreted peptides (activators and repressors) may interact with each other or with receptors. The most highly represented GO category under biological processes was metabolism, associated with 65% of the transcripts, including 21% each for protein and nucleic acid metabolism. Transcripts putatively encoding nuclear or membrane proteins or associated with genes that encode proteins with similar functional domains were also identified. Some transcripts (24% in the cellular component and 16% in molecular category) had unspecified functions. All categories in the GO scheme necessary for eukaryotic development were represented in the transcriptome.

Using KEGG metabolic pathways, 9,411 contigs and singletons were assigned Enzyme Commission (EC) and KEGG Ontology (KO) numbers and represented 11 major metabolic EC pathways (Fig. 2). A similar number of transcripts matched enzymes involved in nucleotide, lipid and energy metabolic pathways, and slightly more were assigned to pathways for amino acid and carbohydrate metabolism (Fig. 2). About 9% of the transcripts were similar to genes involved in xenobiotics, biodegradation and metabolism.

Similarity to ESTs and genes of other nematodes

The *P. zeae* transcripts were further annotated by comparison to sequences of a reference group of 16 nematode species with different modes of feeding and lifestyle. These included three free-living nematodes (FLN): *C. elegans*, *Caenorhabditis remanei* and *Caenorhabditis briggsae*, two animal parasitic nematodes (APN): *A. suum* and *Brugia malayi*, the insect nematode *Pristionchus pacificus*, and ten species of PPNs (three root knot nematodes -*Meloidogyne incognita, Meloidogyne hapla, Meloidogyne javanica*, four cyst nematodes *H. glycines, H. schachtii, Globodera pallida, Globodera rostochiensis*, the migratory endoparasite *R. similis*, the peanut pod nematode *Ditylenchus africanus* and the pine wilt nematode *Bursaphelenchus xylophilus*). Using a TBLASTX search to compare the transcripts to individual sequence databases of the reference nematodes on NCBI at an e-value threshold of 1E-5, a total of 6,473 contigs and 38,181 singletons were similar to sequences of at least one of the nematode species (Fig. 3). For each reference group, there were consistently more singletons with identity

to reference sequences than to contigs. The high number of database sequences for *B. malayi* (compared to *A. suum*) and *H. glycines* (compared to other PPNs) is reflected in the higher number of matches of *P. zeae* transcripts to these species (Fig. 3). Venn diagrams were constructed for the distribution of transcripts among different nematodes, which allocated the transcripts into those that matched sequences of the reference nematodes. Of the annotated transcripts 3,274 contigs and 14,721 singletons were similar to genes/ESTs of nematodes of all lifestyles (Fig. 4). Overall, more transcripts matched ESTs of PPNs than the three well-characterised *Caenorhabditis* spp. There were 1,570 contigs and 15,674 singletons with high similarity to sequences of PPNs, which had no match to any sequence of the FLNs or *P. pacificus:* some of these could represent genes needed for parasitism.

TBLASTX alignment scores (total bit scores) for the annotated transcripts (e-value 1E-5) ranged from 30 to over 4,000. In general, for matches to the nematode species, 40-50% of the annotated contigs had total bit scores of ≥100 whereas more matching singletons (> 60%) had scores of < 100. For example, of the 4,758 transcripts with matches genes of the three *Caenorhabditis* spp., 2,070 had total bit scores ≥ 100 whereas only 15% of the >21, 000 singletons had total bit scores ≥ 100. Unique genes expressed in the transcriptome were determined from the TBLASTX homologues using a total bit score of 100 as the threshold. The results are shown in Fig. 5, which indicates the number of matching genes/ESTs for each of the nematode reference groups and the number of genes/ESTs with unique identifiers for both transcripts. Generally for each reference nematode, more hits/genes were identical to singletons than to contigs.

However, the numbers of ESTs/genes with unique identifiers matching either transcript were similar, indicating that most singletons were fragments of the same genes but lacked enough overlapping sequences to create contigs. To verify this observation, the reads were also assembled with CLC Genomics Workbench 7.5 and GS De novo Assembler (Newbler) Version 2.5 using similar parameters of minimum overlap length and identity of 40% and 90% respectively, and the transcripts compared to *C. elegans* proteins using a local BLASTX with an e-value threshold of 1E-5. The assembly generated 67,849 contigs and 38,960 singletons from the CLC Genomics Workbench and 13, 771 contigs and 13, 938 singletons by Newbler. For both sets of assemblies, some contigs and singletons were identical to transcripts of the same *C. elegans* proteins: 905 singletons of the CLC Genomics Workbench assembly were identical to transcripts of the same proteins matching 2,461 contigs and 190 Newbler singletons and 554 contigs were identical to transcripts of the same *C. elegans* proteins, indicating that for some genes, the read coverage was not sufficient for complete assembly.

For the three *Caenorhabditis* spp., the NextGEne contigs matched 1,354 genes (bit score >100) whereas the singletons matched 1,584 genes, which together represented a total of 2,396 homologues (genes) with unique Wormbase identifiers. Transcripts identical to 1,746 unique *H. glycines* ESTs (with a bit score >100) were identified in the transcriptome - more than for any other nematode species (Fig. 5). The *Caenorhabditis* spp analysis was done on unique genes, whereas for PPNs and APNs the analysis was done with ESTs, and so the number of genes identified may have been over-estimated, because of redundancy in EST databases. Most of the common

genes/ESTs similar to sequences of all the reference nematodes were involved in fundamental biological and molecular processes related to different aspects of development (Table 2).

Comparative analysis with transcriptomes and genomes of PPNs

To extend the depth of the analysis, the transcripts were also compared directly and indirectly (using the Core Eukaryotic Genes used for CEGMA [Core Eukaryotic Genes Mapping Approach]) with the Transcriptome Shotgun Assembly (TSA) of *Pratylenchus thornei* (Nicol et al., 2012), contigs of the *Heterodera avenae* transcriptome (Kumar et al., 2014) and genomic sequences of *M. hapla* (PRJNA29083), *M. incognita* (PRJEA28837), *H. glycines* (PRJNA28939) and *G. pallida* (PRJEB123) available at NCBI. Using a local TBLASTX with a cut-off e-value of 1E-5 with individual sequence databases for each reference nematodes, 9% of the *P. zeae* transcripts (2,631 contigs and 13,349 singletons) were similar to 46% (3,122 of the 6,733) of the *P. thornei* TSA whereas between 20-26% were similar to contigs of the other five PPNs (Fig. 6A). A higher percentage of contigs of *M. hapla* (54%), *M. incognita* (53%) and TSA of *P. thornei* (46%) were similar to *P. zeae* transcripts compared to 36%, 24% and 21% respectively for the cyst nematodes *H. avenae*, *G. pallida* and *H. glycines* (Fig. 6A).

In addition, the 2,748 protein sequences of the euKaryotic clusters of Orthologous Groups (KOGs) representing 458 Core Eukaryotic Genes (CEGs) conserved between *Arabidopsis thaliana*, *C. elegans*, *Drosophila melanogaster*, *Homo sapiens*, *Saccharomyces cerevisiae*, and *Schizosaccharomyces pombe*

(http://korflab.ucdavis.edu/datasets/cegma), were compared to local databases of transcripts and genomes of the seven nematodes above using BLASTX with an e-value threshold of 1E-5 to assess similarity of the genomes to genes expressed in the *P. zeae* transcriptome. Similarities to a total of 454 out of the 458 CEGs were identified in sequences of the seven nematodes (Table S2). For the *P. zeae* transcriptome, 7% of contigs (735) and 2.51% of singletons (3,496) were similar to a total of 431 CEGs (Fig. 6B). A relatively higher percentage of contigs of *M. hapla* (15.3%) and *M. incognita* (12%) were similar to 294 and 364 CEGs respectively. Between 3 – 9% of contigs of the cyst nematodes and 5% of TSA of *P. thornei* were similar to the number of CEGs identified (Fig.6B). In all, 210 CEGs were common to sequences of all the nematode species under study. Transcripts of 23 CEGs were not present in the *P. zeae* transcriptome but appear to be present in the genome of at least one of the other nematodes, and genes for seven of these were present in genomic contigs of root knot and cyst nematodes.

Putative thermosensory, chemosensory and mechanosensory transcripts

C. elegans can detect and respond to temperature changes and gradients as it moves and feeds. Little is known about orthologous genes involved in such activities in PPNs, and for these obligate biotrophs, thermotaxis could play a role in migration through soil and finding a host root. From comparative analysis using 26 *C. elegans* genes with thermosensory and thermotactic functions we found that 12 were similar to 91 *P. zeae* reads (Fig. 6). Among these were *ttx-7* (myo-inositol monophosphatase), two

guanylyl cyclases (*gcy-18, gcy-23*), and two members of the GABA/glycine receptor family of ligand-gated chloride channels (*ggr-1, ggr-2*) (Fujiwara et al. 1996; Kimata et al. 2012). In *C. elegans*, the *gcy-18* and *gcy-23* genes function redundantly to regulate thermotaxis (Hitoshi et al. 2006). Most of these genes are multifunctional and include *tax-4* and *pkc-1* (which encodes a serine/threonine protein kinase) both of which, in addition to controlling several behaviours, are required for sensitivity to chemicals and temperature (thermosensation) (Komatsu et al. 1996; Satterlee et al. 2004).

Nematodes exhibit avoidance or attraction to certain chemicals, but for PPNs relatively little information is available. From our analysis we identified *P. zeae* transcripts similar to genes employed by *C. elegans* to respond to and/or regulate responses to chemical stimuli. *P. zeae* contigs or singletons or both were similar to 44 of 74 *C. elegans* genes with chemosensory functions (Fig. 7). There were also matches to three *C. elegans* genes with 'odorant-response abnormal' phenotypes (*odr-1, odr-2* and *odr-3*) all of which are required for proper functioning of the AWC-olfactory neurons that respond to AWC-sensed odorants (Roayaie et al. 1998; L'Etoile and Bargmann 2000). In addition to *odr-1* (a putative guanylyl cyclase), transcripts similar to three other guanylyl cyclases (*gcy-4, gcy-14* and *daf-11*-a transmembrane guanylyl cyclase) were identified: their functions are not only required for chemosensation and chemotaxis, but results from RNAi in *C. elegans* also indicate expression of *daf-11* may help nematodes avoid volatile and non-volatile odorants (Bargmann et al. 1993; Vowels and Thomas 1994). Most of these genes can nullify effects of some chemicals by modulating neuronal or mechanical movement (e.g. *unc-2*) and are expressed in

chemosensory interneurons. Some putative homologues have roles in avoidance or desensitisation of *C. elegans* to chemicals and/or toxins. An example is the neuropeptide-like protein *nlp-29*, which in *C. elegans* is thought to have anti-microbial properties, since its expression, which is localised in the hypodermis and intestine, is highly up-regulated following exposure to bacteria and fungi (Nathalie et al. 2008).

Sensory receptors are needed by soil-inhabiting nematodes to respond to external stimuli. From comparative analysis with *C. elegans* genes with such roles, 346 *P. zeae* reads (17 contigs and 110 singletons) were similar to 24 *C. elegans* genes that either encode proteins with specific functional domains (e.g. Kunitz, *mec-1*; Kunitz-type protease inhibitor domains, *mec-9*), structural elements (e.g. *mec-7, mec-12*) or regulatory enzymes involved in the complex process of response to touch (Savage et al. 1989). These genes are involved in mobilising the touch receptor 'degenerin' complex (*mec-1, unc-105*) especially to the body wall and in ensuring abundance of collagen in extracellular matrices (García-Añoveros et al. 1998). Amongst these were 6 genes with 'MEChanosensory abnormality' RNAi phenotypes encoding different genes (*mec-1, -2, -7, - 8, -9, -12*), three genes with 'uncoordinated' RNAi phenotypes (*unc-69, -89, -105*) and *gcy-35*.

Guanylyl cyclases may have important functions in several processes, some multifunctional, in protecting nematodes from different forms of abiotic stress. An example is *gcy-35,* which in *C. elegans* expresses in sensory neurons, pharyngeal and body wall muscles. It encodes a soluble guanylyl cyclase, which is required for regulation of feeding and also innate immunity (Gray et al. 2004; Hukema et al. 2006). It

also plays a part in sensing increasing oxygen levels and behavioural changes (Gray et al. 2004). The function(s) of such genes have not been determined in PPNs and this analysis provides a basis for further functional characterisation.

Transcripts similar to C. elegans genes required for feeding and pharyngeal pumping

A total of 386 *P. zeae* reads were similar to 12 *C. elegans* genes required for pharyngeal pumping and feeding. Of these, 214 were identical to *eat-3*, *eat-4*, *eat-5* and *eat-6* for which loss of function affects 'eating'. These genes encode proteins involved in different processes in *C. elegans,* but are generally expressed in pharyngeal muscles and/or cells (Avery 1993; Starich et al. 1996). An example is e*at-3* which encodes a mitochondrial dynamin required for body size development, growth, movement and reproduction. It is highly expressed in muscles and neurons, and loss of function in *C. elegans* results in irregular and reduced pharyngeal pumping, as do transcripts of *eat-4* (a vesicular glutatmate transporter), *eat-5* (an innexin) and *eat-6* (encoding an alpha subunit of a sodium/potassium ATPase) (Kanazawa et al. 2008). Loss of function also affects other aspects of feeding: e*at-4* is required for chemotaxis, feeding, foraging and thermotaxis, whereas *eat-5* and *eat-6* are required for relaxation and synchronised pharyngeal muscle contractions (Starich et al. 1996; Lee et al. 1999). The roles of similar genes have not been studied in PPNs, for example how they may affect host recognition and feeding. *P. zeae* transcripts similar to *C. elegans* genes encoding products which regulate action potentials of pharyngeal muscles or are expressed in pharyngeal muscles or neurons were a synaptotagmin (*snt-1),* and neuropeptide-like proteins including an

LQFamide neuropeptide (*nlp-12*) possibly with a role in the regulation of digestive enzyme secretion and fat storage, an MSFamide neuropeptide (*nlp-13*) and five FMRFamide-like peptides (*flp-5, flp-8, flp-14, flp-16* and *flp-18*): orthologues of these genes have also been identified in other PPNs (Abad et al. 2008; Maule and Curtis 2011).

P. zeae transcripts matching genes for locomotion

Uncoordinated (*unc*) mutants of *C. elegans* have been used to characterise over 700 genes involved in many cellular processes that regulate mechanical and neural control of movement, including myosin assembly (Hoppe and Waterston 2000), regulation of G protein and receptor signalling (Koelle and Horvitz 1996; Hajdu-Cronin et al. 1999; Lackner et al. 1999; Robatzek et al. 2001), neuropeptide function (Nelson et al. 1998; Frooninckx 2012) and maintenance of structure and integrity of the cuticle (Broday et al. 2007). These genes were used to identify *P. zeae* transcripts that may encode proteins with similar functions. About 1,300 reads had high similarity to *C. elegans* genes involved in structural integrity of the cuticle, muscle and collagen formation and function, and directional movement (Fig. 7).

Amongst matches to genes of *C. elegans* were those encoding nematode cuticular collagen (e.g. *col-2, rol-6*) required for development of normal body morphology and those required for structural integrity, proper functioning and maintenance of muscle including the troponin C and I (*tnc*, *tni*) gene families (Terami et al. 1999). There were 7,710 reads matching the latter genes, including *paralysed arrest at two-fold* (*pat-2, -3, -4, 6-, 10, -12)*, genes encoding heavy chain myosin (e.g. *myo-1*),

ttn-1, and *vab-10*. Of these, *pat-10* (alias *tnc-2*) with *unc-87* (required to maintain structure of myofilaments in body wall muscle cells) are also involved in control of movement in RLNs: knockdown of both genes causes paralysis and strong inhibition of reproduction in host tissues (Soumi et al. 2012; Tan et al. 2013). A total of 24 contigs and 10 singletons matched *ttn-1* (a connectin), the largest known protein responsible for passive elasticity of muscle (Forbes et al. 2010). Transcripts similar to the *vab-10* gene were highly represented in the transcriptome: there were 2,263 reads assembled into 17 contigs and 39 singletons. In *C. elegans*, *vab-10* encodes two spliceoforms of spectraplakins that act jointly to provide mechanical resilience to the epidermis under strain. RNAi phenotypes of *vab-10* show that it is essential for organisation of body wall muscles (Plenefisch et al. 2000; Bosher et al. 2003).

Transcripts with high identity to two functionally diverse neuropeptides, f*lp-1* and *flp-7* of *C. elegans* were also identified. The *flp-1* gene regulates well-coordinated sinusoidal movement and transition between active and inactive states of egg laying (Nelson et al. 1998). In contrast *flp-7* is a negative regulator of movement, since injection of *C. elegans flp-7* peptides into *A. suum* causes paralysis and loss of locomotory waveforms (Reinitz et al. 2000). In addition, 162 contigs matched 84 '*unc*' genes encoding proteins involved processes for which loss-of-function results in aberrant movement or uncoordinated phenotypes. As well as the transcripts putatively encoding *slo-1* and *sdn-1* genes, which are required for forward and backward motion in *C. elegans,* functional characterisation of similar genes which could be important for mechanical and neural control of movement of migratory endoparasitic nematodes is

needed since such genes are potential targets for control of these pests (Carre-Pierrat et al. 2006).

Nematode parasitism genes

There were 92 contigs and 674 singletons identified with high sequence similarity to 25 previously characterised or putative PPN parasitism genes, or those that encode proteins involved in nematode-host interactions (Table 3). Seven of these transcripts matched only singletons and five were similar to gene encoding products employed by PPNs to elicit or suppress host defences: these included secreted SPRYSEC proteins of *G. pallida* (Sacco et al., 2009), superoxide dismutase (Zacheo et al., 1987, Zacheo and Bleve Zacheo, 1988), thioredoxin (Hewitson et al., 2008; Lu et al., 1998), peroxiredoxin (Li et al., 2011), glutathione peroxidase (Jones et al., 2004) and glutathione S-transferase (Debreuil et al., 2007). The *P. zeae* sequences were most similar to genes of three *Meloidogyne* spp, two *Globodera* spp, two *Heterodera* spp and *D. africanus,* and except for peroxiredoxin, were also similar to genes of the roundworm, *A. suum* (Table 3). A candidate homologue of the nematode-specific fatty-acid- and retinol-binding (FAR) family of proteins, secreted for evasion of host defences, was also identified with high total alignment scores to those characterised for other PPNs. Many transcripts were identical to three well-characterised cell-wall degrading enzymes apparently secreted by sedentary PPNs during tissue migration: pectate lyase, endoglucanases and polygalacturonase (Table 3). Transcripts were also identified that matched three groups of proteases secreted by nematodes of all lifestyles: these were

aminopeptidase, serine protease and cathepsins B, D, S and L types of proteases. RLNs secrete a range of proteins: some are involved in digestion and others in developmental processes (e.g. moulting), others are secreted into host cells and may be required for parasitism (e.g. influencing signalling pathways, evading host defences). Some transcripts matched genes with such functions and these included the secreted 14-3-3/b protein and calreticulin of *M. incognita*, and annexins which also had high similarities to those of cyst nematodes (Jaubert et al., 2004, 2005, Table 3). A total of 14 contigs and 107 singletons matched the group of PPN-specific transthyretin-like proteins (*ttl*) and precursors, with the best matches to those of *R. similis*, for which *ttl 1-4* are preferentially expressed in the parasitic stages (Jacob et al. 2007).

A notable feature was that many transcripts similar to those encoding effectors and thought to be required for giant cell or syncytium formation by root knot and cyst nematodes were not found in the *P. zeae* transcriptome. Examples of these include the secreted effector 7E12 (AF531166.1), for which over-expression in tobacco accelerates giant cell formation induced by *M. incognita* (de Souza Junior et al., 2011), and CLE peptide and 16D10 CLE related proteins (e.g. DQ087264.1) known to influence transcriptional regulation processes that promote giant cell induction (Huang et al., 2006). Similarly, we did not find homologues of the C-terminally Encoded Peptide (CEP) family of regulated peptides which contribute to control of root development in vascular plants (Matsubayashi, 2011). A recent survey identified similar peptides in the *M. hapla* genome. These peptides have structural similarities to those present in plants, and have been suggested to play a role in modifying root structures during the

nematode infection (Bobay et al., 2013). In a local BLASTX comparison of the *P. zeae* transcripts to CEPs of *M. hapla* (*MhCEP1-12;* Bobay et al., 2013), plant homologues of *Medicago truncatula* (*MtCEP1-11;* Bobay et al., 2013) and those in *Arabidopsis thaliana* (*AtCEP1-5*; Bobay et al., 2013), none of the transcripts were similar to any of the CEPs at an e-value 1E-3 threshold. In addition, we did not find any *P. zeae* transcript matching the taxonomically restricted MAP-1 gene (e.g. AJ278663.1) of *Meloidogyne* species, which is potentially involved in the early stages of host recognition (Semblat et al., 2001). In the same way, we did not idenitify any transcript similar to three cyst nematode genes which encode effectors secreted via the oesophageal gland cells. These were the effectors *Hg30C02 effector* (JF896103) of *H. glycines,* expression of which increases susceptibility of Arabidopsis to infection by *H. schachtii*, *Hs19C07* (AF490250.2) which interacts with the Arabidopsis auxin influx transporter LAX3 to facilitate syncytium development, and *10A06* (GQ373257.1) of *H. schachtii*: host expression of this gene induces morphological changes, targets spermidine synthase and possibly disrupts Arabidopsis defense signalling resulting in increased susceptibility to infection (Hewezi et al., 2010; Lee et al., 2011; Hamamouch et al., 2012).

Putative secretome of P. zeae.

The set of proteins identified in PPN secretions which appear to play specific roles in host interactions are collectively referred to as the secretome. For sedentary endoparasites, such proteins have typically been isolated from gland cells, whereas for the filarial nematode *B. malayi*, they have been identified from excretory-secretory

products of microfilariae, adult males and females at the host/parasite interface. A TBLASTX comparison of the *P. zeae* transcriptome with the secretomes of *H. glycines*, *M. incognita* and *B. malayi* revealed 193 contigs and 1,233 singletons with sequence similarity to genes of 61 secreted peptides (Table 4). These were divided into 10 groups based on function, including those involved in structural and cytoskeletal functions, energy metabolism, and protein digestion and fate (Table 4). There were also matches to genes of peptides secreted in response to stress, those suggested to be involved in evasion of host immune responses, and in the case of the PPNs, proteins secreted to modify plant cell wall polysaccharides. Amongst the 20 secreted peptides common in secretomes of *M. incognita* and *B. malayi* were structural proteins (e.g. actin, tropomyosin family protein and high mobility group proteins), ubiquitination proteins, peptidases (e.g. aminopeptidase and serine carboxypeptidase), and proteins of unknown functions e.g. transthyretrin-like proteins. Transcripts with high percentage identity to those encoding six proteins identified in three independent studies of excretory-secretory products of *B. malayi* life stages were also identified. Four of these (glutathione peroxidase, superoxide dismutase, galectin, and cystatin-type proteinase inhibitor) were also identified by proteomic analysis of the secretome of *M. incognita* (Bellafiore et al. 2008). The other two were the macrophage migration inhibitory factor 1, also an encoded protein in *C. elegans*, and triose-phosphate isomerase (Bennuru et al. 2009). Thirteen secreted products of *M. incognita* and *H. glycines*, whose genes share high homologies to transcripts of *P. zeae*, have so far not been identified in excretory-

secretory products of *B. malayi*, these were 8 of the parasitism genes in Table 3, four structural proteins and an ATP synthase (*atp-2*).

In addition, 765 *P. zeae* transcripts matched 531 ESTs isolated from the single dorsal and/or the two subventral oesophageal gland cells in pre-parasitic and parasitic stages of *H. glycines* (Wang et al. 2001; Gao et al. 2003). Characterisation of these ESTs and those identified above is needed to determine if they are indeed secreted by *P. zeae*.

Functional characterisation of P. zeae transcripts

P. zeae transcripts putatively encoding five proteins were characterised using *in situ* hybridisation and/or RNA interference: these were putative transcripts for troponin C (with the *C. elegans* RNAi phenotype *pat-10*) and calponin (*unc-87*) both of which are essential for structural integrity and proper muscle contraction in nematodes, and three putative parasitism genes, β-1, 4-endoglucanase (*eng*), calreticulin (*crt-1*) and a transcript for the SXP-RAL2 protein (*sxp*). Primers were designed using transcripts with highest percentage identity to characterised genes and were used successfully to amplify *P. zeae* transcripts from mRNA used for the transcriptome sequencing using reverse transcription-PCR as described in the Supplementary Methods, S1: the amplicons were designated *Pzpat-10* (282 bp), *Pzunc-87* (275 bp), *Pzeng* (400 bp), *Pzcrt-1* (354 bp) and *Pzsxp* (234 bp) (Supplementary Methods, S1). Sequences of the amplicons were determined by Sanger sequencing, and each was 100% similar to the contigs/singletons from which primers were designed (Supplementary Methods, S1,

Table S3a-e). TBLASTX search revealed they were also identical to genes of many PPNs (Table S3a-e). Notably, matches to *Pzeng* included endoglucanases with GH5 activity of several bacteria including species of *Sorangium*, *Saccharophagus*, *Cellulophaga* and *Hymenobacter* suggesting acquisition of such genes from bacteria via horizontal gene transfer (Table S3a).

The *P. zeae* amplicons were then cloned into a transcription vector, pDoubler: we designed this vector for synthesising RNA probes and dsRNA for *in vitro* RNA interference (Supplementary Methods, S1). DIG-labelled RNA antisense probes synthesised from DNA templates digested from pDoubler clones for *Pzeng, Pzcrt-1 and Pzsxp* all hybridised to the pharyngeal region of *P. zeae* juvenile and adult stages, behind the median bulb suggesting that they may be secreted via the pharyngeal glands (Fig. 8-I). In contrast, for *Pzpat-10* and *Pzunc-87,* intense non-specific hybridisation was observed in all stages of the nematodes and hybridisation was generally localised at the anterior end of the nematodes (Fig. 8-I). There was no hybridisation with sense probes of any of the genes.

To further characterise the three putative parasitism genes using RNAi, dsRNA corresponding to *Pzeng, Pzcrt-1, Pzsxp* and to a 715 bp of the green fluorescent protein (GFP) were synthesised using the transcription vector pDoubler which has two opposing T7 polymerase promoters on either side of a multiple cloning site as described in Supplementary Methods S1. For each gene, 2,000 vermiform *P. zeae* were soaked in 2 µg/µL dsRNA, (designated dsPzeng, dsPzcrt-1, dsPzsxp *and* dsGFP) resuspended in M9 buffer with 0.05% gelatin, 3 mM spermidine trihydrochloride and 50 mM octopamine

and incubated for 16 h at 25°C. Compared to nematodes soaked in the medium without dsRNA and with dsGFP, both used as controls, no effect was observed on the behaviour and activity of *P. zeae* nematodes soaked in dsRNA of the endogenous genes for 16 h. However, when transcript abundance of the endogenous target genes in 500 of these nematodes was assessed 16 h after incubation by quantitative PCR, and relative expression determined using the ΔΔCt with 18S rRNA expression as internal control, there was a reduction of expression for all three genes. Target gene expression levels were determined for control nematode treatments (either no dsRNA or treated with dsGFP) and dsRNA of target genes: this enabled identification of the effects of the stress of feeding a non-specific dsRNA from the effects of down-regulating expression of the target genes. There was a 207 fold reduction in transcript abundance for *Pzeng,* 67 fold for *Pzcrt-1* and 36 fold reduction for *Pz-sxp* compared to expression in nematodes soaked with dsGFP (Fig. 8-II).

To assess possible effects of down-regulation of target genes by RNAi on nematode reproduction, 50 nematodes each from the *in vitro* RNAi soaking treatments were used to infect five replicates of carrot mini discs and the number of nematodes counted after 5 weeks, as described by Tan et al. (2013). After this period, the number of nematodes present on carrot discs not treated with dsRNA or with dsGFP increased about 9 times and there was no significant difference between the two treatments ($p<0.05$, Fig. 8-III). For the nematodes treated with dsRNA to target genes, the numbers extracted after 5 weeks for those soaked in *dsPzcrt-1* number increased 7 times over the primary inoculum, but this did not differ significantly from the controls ($p<0.05$, Fig. 8-

III). In contrast, there was a significant reduction in reproduction of nematodes soaked with dsPzsxp and dsPzeng: after 5-weeks culture the number of nematodes soaked in dsRNA of these targets was substantially less, with an 84% reduction in numbers for dsPzeng compared to the control (no dsRNA) (Fig. 8-III). These results mirror previous work in which we showed that similar treatment of *P. zeae* with dsRNA of *Pzpat10* and *Pzunc87*, both of which substantially reduced transcript levels and significantly reduced reproduction of *P. zeae* in carrot discs (Tan et al., 2013).

Discussion

This study provides the first sequence data on the transcriptome of the root lesion nematode *P. zeae*, a significant pest of crop plants such as sugarcane and sorghum. Analysis of the data focused on transcripts likely to encode structural proteins, those involved in locomotion, sensitivity to external stimuli and for host parasitism. About 60% of the total reads were assembled with SoftGenetics NextGENe V2.16 software into 10,163 contigs leaving 139,104 singletons. This relatively high number of singletons could have been a result of stringent base overlapping parameters used for contig assembly, which could also lead to singletons matching the same BLAST hits as contigs (and in some cases, more singletons having higher BLAST hit percentage identities to hits than contigs). However, assembly of the reads with both CLC Genomics Workbench and Newbler also generated singletons, although less than in the NextGEne assembly. A BLASTX of the resulting transcripts with *C. elegans* proteins showed some singletons and contigs from each assembly were identical to the same *C. elegans*

proteins. This suggests that the read coverage may not be large enough for a complete assembly. Nevertheless, the use of both contigs and singletons of appropriate length in annotation, as done here, provides useful new information, because some singletons not only matched similar characterised genes, but had high enough percentage identities to warrant further characterisation, as demonstrated by the strong reduction in nematode reproduction on carrot discs after soaking *P. zeae* with dsRNA of *Pzcrt-1* and *Pzsxp*. About 65% of contigs and 28% of singletons of the NextGEne assembly were similar to sequences of different nematodes (FLNs, APNs and PPNs). The transcripts matching genes common to all these nematode species mainly encoded structural proteins or were involved in general metabolic and biological processes.

For RLNs like *P. zeae* all juvenile stages (except J1s) and adults, can move to and feed from a range of cells at the epidermis and within roots. In contrast, sedentary endoparasitic nematodes enter and migrate in host roots, and identify one or more specific host cells at which they control cell differentiation to develop permanent feeding sites (different forms of giant cells or syncytia, Jones 1981; Jones and Goto, 2011) after which the nematodes become sedentary. Since all feeding stages of *P. zeae* are motile and active it is not surprising that most genes/transcripts are likely to be expressed throughout the infective stages of their life cycle. One such group of genes are those needed for different forms of movement, and *P. zeae* transcripts similar to 84 *C. elegans* genes were identified in the transcriptome with uncoordinated RNAi phenotypes (e.g. *unc-87*), *slo-1* and *sdn-1* required for directional movement and two neuropeptides (*flp-1* and *flp-7*) that regulate sinusoidal movement and locomotory

waveforms (Nelson et al. 1998; Reinitz et al. 2000). The importance of proper functioning of two such genes (*pat-10* and *unc-87)* has been demonstrated clearly in RNAi soaking experiments for three *Pratylenchus* species including *P. zeae,* in which these two genes were down-regulated after soaking in dsRNA and subsequently there was a significant reduction in reproduction (Soumi et al. 2012; Tan et al. 2013).

One interesting aspect of this study was the identification of many *P. zeae* transcripts with similarity to genes involved in mechanosensory, chemosensory and thermosensory activities employed by *C. elegans* in response to biotic and abiotic stresses. There is no information available on whether *P. zeae* employs similar genes and mechanisms to sense and respond to chemicals in the rhizosphere, or to mechanical pressures in the soil environment, to temperature changes, or what mechanisms it uses to evade or defend itself against plant immune responses as such genes have been little studied in PPNs. Similarly, the free-living *C. elegans* can be infected by pathogenic viruses and bacteria. Again, little is known about interaction of PPNs and potential pathogens, although the 'Soybean cyst nematode midway virus' was found recently in a transcriptome of *H. glycines* (Bekal et al. 2011). The identification of transcripts similar to the *C. elegans* genes *nlp-29*, *bre-3*, *bre-4* and *bre-5* in the *P. zeae* transcriptome is therefore interesting and warrants further investigation - in *C. elegans* their roles are to protect the nematode from biotic stresses.

This study adds to the increasing molecular information available for RLNs, and highlights the similarities and differences in gene content between migratory and sedentary endoparasites (Jones and Northcote 1972; Jones 1981; Jones and Fosu-

Nyarko 2014). Our analysis shows that *P. zeae* has 19 putative parasitism genes similar to those of *P. coffeae* and *P. thornei* (Haegeman et al. 2011; Nicol et al. 2012). However, transcripts of seven effectors widely reported to be important in sedentary nematode-host interactions and identified in the *P. zeae* transcriptome were not reported as expressed in the transcriptomes of *P. coffeae* and *P. thornei* (Haegeman et al. 2011; Nicol et al. 2012). These include transcripts for superoxide dismutase (required for stress response and protection), two proteases (aminopeptidase and serine protease) and *SKP-1,* which has been identified in secretions of *M. incognita* and *H. glycines* and is thought to be involved in ubiquitination (Gao et al. 2003; Bellafiore et al. 2008). Interestingly, our data, and similar data from *P. coffeae* and *P. thornei,* highlight at the molecular level differences in the lifestyles between migratory and sedentary endoparasitic nematodes. It further confirms that some genes present in sedentary endoparasitic nematodes thought to be essential for host-interactions have not been identified so far in migratory RLNs. The genes apparently missing include secreted effectors involved in giant cell or syncytium formation identified in sedentary endoparasitic nematodes such as 7E12 (de Souza Junior et al., 2011), C-terminally Encoded peptides (Bobay et al., 2013), CLE peptide and 16D10 CLE (Huang et al., 2006) and MAP-1, required for host recognition by root knot nematodes (Semblat et al. 2001) and effectors 10A06, Hs19CO7 and 30C02 of cyst nematodes (Hewezi et al., 2010; Lee et al., 2011; Hamamouch et al., 2012). In the transcriptome analyses of *P. thornei* and *P. coffeae* (Nicol et al., 2012, Haegeman et al., 2011), transcripts were reported with very low percentage identity to the chorismate mutase encoded by sedentary nematodes,

but no such match was found in the *P. zeae* transcriptome. It is not clear whether *Pratylenchus* spp require this gene for host interactions, but based on the characterised functions of chorismate mutases of fungi, plants and those of sedentary endoparasitic nematodes, it seems probable that a functional chorismate mutase is not present in RLN genomes.

Despite different lifestyles, there is nevertheless a common set of proteins or effectors in PPNs which enable them to invade root tissues and evade or suppress host immune responses. From this analysis, *P. zeae* also expresses transcripts that may encode 26 such effectors previously identified in sedentary PPNs and in other RLNs. These are required for common parasitic processes, including cell wall modification (pectate lyases and endoglucanases) required for migration in plant tissues. There is also a similarly high number of transcripts putatively encoding proteases, and gene products required to combat stresses, such as glutathione S-transferase, superoxide dismutase and reducing agents such as peroxiredoxin and thioredoxin that may act against reactive oxygen species generated from the host in addition to involvement in nematode metabolism. There have not been any proteomic studies published so far on analysis of proteins secreted by *Pratylenchus* spp. However, we have identified a number of transcripts like those that encode similar proteins secreted by *B. malayi*, *M. incognita* and *H. glycines,* which suggests that parasitic nematodes may have similar genes, but the functions of these need to be determined for each nematode, because they may have evolved divergent roles that reflect different evolutionary pressures or lifestyles.

Although understanding of the molecular basis of host-pathogen interactions of RLNs has been less studied than that for sedentary PPNs, new sequencing technologies now make it possible to sequence whole transcriptomes or genomes of other PPNs such as economically important *Pratylenchus* species. Here we provide new data on preliminary characterisation of many expressed transcripts of *P. zeae*, including many potentially involved in nematode development, migration and feeding from cells of host tissues and some which appear to be involved in protecting the nematode in soil and to counteract host defences in plant tissues. In this study, we have also examined the spatial expression of transcripts encoding putative homologues of five proteins required for movement or nematode parasitism by *in situ* hybridisation. This analysis demonstrated that transcripts thought to encode a β-1,4-endoglucanase gene (*Pzeng*)and calreticulin (*Pzcrt-1*) and a putative transcript of a SPRY containing protein (*Pzsxp*) are localised to the *P. zea*e pharyngeal gland cells, which suggest that they may be secreted. Significantly, knockdown of *Pzeng* and *Pzsxp* by RNAi significantly reduce multiplication of *P. zeae* in carrot discs. Although most of the genes identified need functional characterisation, the data generated here has enabled identification of potential targets for developing new control strategies (Soumi et al. 2012; Tan et al. 2013; Jones and Fosu-Nyarko, 2014; Fosu-Nyarko and Jones, 2015). In addition, the characterisation of transcripts with unknown function provides a further challenge: functional characterisation is now required to generate a more complete picture to complement future sequencing and assembly of the genomes of *Pratylenchus* species.

Experimental Procedures

Nematodes and RNA extraction

Vermiform juveniles and adults of *P. zeae* were harvested from roots of infected sugarcane using a mist apparatus (Tan et al., 2013). Nematodes which migrated through the coffee filter were collected via tubing in a disposable syringe. Active nematodes were removed from the syringe at 4 hr intervals, and pelleted by gentle centrifugation. The nematodes were then resuspended in sterile water and washed by centrifugation and re-suspension a further five times. The nematodes were then surface sterilised and inoculated onto surface sterilised carrot pieces maintained *in vitro* (Tan et al 2013). Nematode surface sterilisation involved suspending nematodes in 1% chlorhexidine gluconate (hibitane) for 20 min, followed by 1% streptomycin sulphate for 5 min: they were then washed five times with sterile water. Washing involved suspending nematodes in sterile water and centrifuging them at 1,000 g for 3 min: re-suspension was by gently inverting the tubes 3-5 times; repeated 5 times. The nematodes were then examined by light microscopy to check for viability and any signs of microbial contamination before use. To further minimise any possible contamination before RNA extraction for transcriptome sequencing, mixed juvenile and adult stages of *P. zeae* were extracted from the sterile carrot pieces and surface-sterilised again. Nematode inocula added to carrot mini disc cultures and for RNAi feeding experiments were obtained from the axenic stock cultures without further sterilisation, after extraction and washing with sterile distilled water.

%%Total RNA was extracted from freshly collected active nematodes using Trizol (Invitrogen Life Technologies, Carlsbad, USA) and the RNA cleaned with the RNeasy Mini kit (Qiagen Inc, California, USA). The quality and quantity of the RNA was assessed with an Agilent 2100 Bioanalyzer (Agilent Technologies, Mississauga, Canada): RNA with RNA Integrity Number (RIN) greater than 7 with absorbance ratios of 260:280 at 2.1 and 260:280 at 2.0 was used for cDNA synthesis.

Sequencing, contig assembly

A cDNA library was prepared from 2.5 µg of total RNA using the Ovation RNA-Seq system (NuGEN Technologies, Inc., CA, USA) and the quality assessed using an Agilent 2100 Bioanalyzer. Half a picotitre plate was used to generate the 454 GS FLX dataset using a Roche 454 GS FLX at the Institute of Immunology and Infectious Diseases, Murdoch University. Sequencing linkers were removed from all reads and *de novo* assembly was carried out using standard settings of the SoftGenetics NextGENe V2.16 software and the Condensation Tool (Manion et al. 2009). The average PHRED score of the sequences were determined using the CLC Genomics Workbench 7.5. The reads were also assembled with the CLC Genomics Workbench 7.5 and GS De novo Assembler (Newbler) Version 2.5: the purpose was to assess if the presence of singletons after the NextGEne assembly was a result of the algorithms of the software or whether it related to read coverage. The set of parameters used for the CLC Genomics Workbench assembly were: mismatch cost of 2, insertion cost of 2, deletion cost of 2, length fraction of 0.4 and similarity fraction of 0.9. The Newbler assembly was done with the following

settings: a seed step of 12, seed length of 16, seed count of 1, minimum overlap length of 40%, minimum overlap identity of 90% and alignment identity and difference scores of 2 and -3 respectively.

Calculation of Alien Index (AI) for transcripts

The detection of possible contaminating sequences in the transcriptome was assessed using the Alien Index (AI) analysis. AI was calculated for *P. zeae* transcripts with a BLAST hit to at least one metazoan or a non-metazoan sequence in the NCBI non-redundant nr/nt database using the following formula described by Gladyshev et al. (2008) and Eves-van den Akker et al. (2014): AI=log(best metazoan e-value + 1E−200) − log(best non-metazoan e-value + 1E−200). When no metazoan or non-metazoan significant BLAST hit was found, an e-value of 1 was automatically assigned, consequently, no AI value was obtained for transcripts where there was no significant hit in the non-redundant database. Transcripts with an AI >0 indicates a better hit to a non-metazoan species than to a metazoan species, were considered possible contaminants and excluded from further analysis.

Functional classification

The transcripts were characterised using TBLASTX with a cut-off expected (E) value of 1E-5 (Camacho et al. 2009; Hunter et al. 2009). The Gene Ontology (GO) functional classification scheme was used to assign the transcripts into three GO functional groups: molecular function, biological processes and cellular components

(Ashburner et al. 2000). The molecular function category described the putative functions of transcripts based on their involvement in fundamental and cellular activities, whereas the biological process category included transcripts encoding gene products involved in molecular events and pathways. The cellular component category grouped the transcripts that encoded structural proteins of cells or its extracellular environment. As an alternative approach to annotation and to determine their biochemical functions, the transcripts were assigned to metabolic pathways using the Kyoto Encyclopaedia of Genes and Genomes (KEGG, http://www.genome.jp/KEGG/) enzyme commission (EC) and orthology (KO) pathways (Kanehisa et al. 2010). TBLASTX results against KEGG databases (nucleotide and amino acid) were parsed to obtain KEGG identifiers and EC numbers which were matched to specific metabolic pathways. The results were then compiled and percentages of the transcripts in each functional group (or pathway) determined.

Comparative analysis with nematode sequences

Transcripts were compared to sequences of a reference group of nematodes of four different lifestyles (free-living, insect, animal and PPN) in the public databases of National Centre for Biotechnology Information (NCBI) Genbank (www.ncbi.nlm.nih.gov), Uniprot (www.uniprot.org) and Wormbase (www.wormbase.org) using the TBLASTX. The comparative analysis used nucleotide and protein sequences of three annotated free living nematodes *C. elegans, C. remanei* and *C. brigssae*, the symbiotic insect nematode *P. pacificus*, and two species of animal parasitic nematodes: the roundworm *A. suum* and the filarial worm *B. malayi*. Most comparisons with PPNs used expressed

sequence tags (ESTs) and the full-length cDNA sequences of three root knot nematodes (*Meloidogyne* spp), four cyst nematodes (*Heterodera* and *Globodera* spp), the migratory nematode *R. similis*, pine wilt nematode *B. xylophilus* and the peanut pod nematode *D. africanus* available in the *nr* and EST databases of NCBI. The TBLASTX process was scripted to retrieve, amongst others, maximum and total bit/alignment scores of transcripts to matches from the databases with a cut-off e-value of 1E-5. The retrieved XML files were converted to csv formats for easy access and analysis. Because of redundancy in public databases some transcripts had matches to different sequences. Also different regions of some contigs matched different ESTs or genes with a range of total bit scores. For these, the contig with the highest total bit score was considered most similar to a hit. Detailed analysis and comparisons (nucleotide and protein domain alignments) of some transcripts to TBLASTX matches were translated and the amino acid aligned to confirm identity to matches using BioEdit (Hall 1999), ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/)and/or the suite of programmes available at Expasy (www.expasy.org).

The transcripts were also compared with transcriptomes of *Pratylenchus thornei* and *H. avenae* and publicly available genomic contigs of *M. hapla* (PRJNA29083), *M. incognita* (PRJEA28837), *H. glycines* (PRJNA28939) and *G. pallida* (PRJEB123) by direct TBLASTX matching with the reference database and indirectly through BLASTX identification of genes similar to the proteins encoded by the Core Eukaryotic Genes (CEGs) used in Core Eukaryotic Genes Mapping Approach, CEGMA (Parra et al., 2007).

Annotation of transcripts for specific functions and processes

Transcripts with carbohydrate enzyme activity were identified using the Carbohydrate active enzyme database (CAZy, www.cazy.org). In addition, detailed functional analysis of the transcriptome was done to identify transcripts encoding proteins potentially required for parasitism. All transcripts that matched genes and ESTs of previously characterised or putative parasitism genes of PPNs identified from the TBLASTX search of the nr/nt database were aligned directly with the matched nematode sequences and the best matches recorded as in Table 3. The extensive resource at Wormbase (www.wormbase.org) was primarily used to identify *C. elegans* homologues required for locomotion, eating and pharyngeal regulation, and those that might protect the nematode against biotic and abiotic stresses. Transcripts putatively encoding proteins involved in chemoreception and chemotaxis, thermo- and mechanosensory reception were also identified using sequences of *C. elegans* from Wormbase (sequence release WS238).

Functional characterisation of transcripts

Transcripts putatively encoding *pat-10 (Pzpat-10)*, unc-87 (*Pzunc-87*), β-1,4-endoglucanase (*Pzeng*), calreticulin (*Pzcrt-1*) and SXP-RAL2 (*Pzsxp*) were functionally characterised using *in situ* hybridisation and/or RNA interference. Dig-labelled probes for sense and antisense fragments for each of the genes, amplified from cDNA of mixed stage *P. zeae*, were generated and hybridisation using a modified protocol of de Boer *et al*. (1998). RNAi of the three putative parasitism genes, *Pzeng*, *Pzcrt-1* and *Pzsxp,* were

studied by soaking mixed stage nematodes with dsRNA corresponding to the genes and *gfp* for 16 h after which gene knockdown and its effect on reproduction on carrot mini discs were assessed (Tan et al. 2013). Detailed methodology for the functional characterisation is supplied in the Supplementary Methods and Table S3a-e.

Acknowledgements

We thank the Australia India Strategic Research Fund (Grant BF030027), Nemgenix Pty Ltd, and Murdoch University for financial support.

References

Abad, P., Gouzy, J., Aury, J.-M., Castagnone-Sereno, P., Danchin, E. G. J., Deleury, E., Perfus-Barbeoch, L., Anthouard, V., Artiguenave, F., Blok, V. C., Caillaud, M.-C., Coutinho, P. M., Dasilva, C., De Luca, F., Deau, F., Esquibet, M., Flutre, T., Goldstone, J. V., Hamamouch, N., Hewezi, T., Jaillon, O., Jubin, C., Leonetti, P., Magliano, M., Maier, T. R., Markov, G. V., McVeigh, P., Pesole, G., Poulain, J., Robinson-Rechavi, M., Sallet, E., Segurens, B., Steinbach, D., Tytgat, T., Ugarte, E., van Ghelder, C., Veronico, P., Baum, T. J., Blaxter, M., Bleve-Zacheo, T., Davis, E. L., Ewbank, J. J., Favery, B., Grenier, E., Henrissat, B., Jones, J. T., Laudet, V., Maule, A. G., Quesneville, H., Rosso, M.-N., Schiex, T., Smant, G., Weissenbach, J. and Wincker, P. (2008) Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. *Nat. Biotechnol.* **26,** 909-915.

Adam, M.A.M., Phillips, M.S., Jones, J.T., Blok, V.C. (2008) Characterisation of the cellulose-binding protein Mj-cbp-1 of the root knot nematode. *Physiol Mol Plant Pathol* **72**:21–28.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, M. J., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M. and Sherlock, G. (2000) Gene Ontology: tool for the unification of biology. *Nat. Genet.* **23,** 25-29.

Avery, L. (1993) The genetics of feeding in *Caenorhabditis elegans*. *Genetics* **133,** 897-917.

Bargmann, C. I., Hartwieg, E. and Horvitz, H. R. (1993) Odorant-selective genes and neurons mediate olfaction in *C. elegans*. *Cell* **74,** 515-527.

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.

Bennuru, S., Semnani, R., Meng, Z., Ribeiro, J. M. C., Veenstra, T. D. and Nutman, T. B. (2009) *Brugia malayi* excreted/secreted proteins at the host/parasite interface: stage- and gender-specific proteomic profiling. *PLoS Negl. Trop. Dis.* **3**.

Blair, B. L. and Stirling, G. R. (2007) The role of plant-parasitic nematodes in reducing yield of sugarcane in fine-textured soils in Queensland, Australia. *Aust. J. Exp. Agr.* **47,** 620-634.

Bobay, B. G., DiGennaro, P., Scholl, E., Imin, N., Djordjevic, M. A., and Bird, D. M. (2013). Solution NMR studies of the plant peptide hormone CEP inform function. *FEBS letters*, **587**, 3979-3985.

Bosher, J. M., Hahn, B.-S., Legouis, R., Sookhareea, S., Weimer, R. M., Gansmuller, A., Chisholm, A. D., Rose, A. M., Bessereau, J.-L. and Labouesse, M. (2003) The *Caenorhabditis elegans vab-10* spectraplakin isoforms protect the epidermis against internal and external forces. *J. Cell Biol.* **161,** 757-768.

Broday, L., Hauser, C. A., Kolotuev, I. and Ronai, Z. (2007) Muscle-epidermis interactions affect exoskeleton patterning in *Caenorhabditis elegans*. *Dev. Dynam.* **236,** 3129-3136.

Camacho, C., Coulouris, G., Avagyan, V., Ning, M., Jason, P., Kevin, B. and L., M. T. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10,** 421.

Carre-Pierrat, M., Grisoni, K., Gieseler, K., Mariol, M. C., Martin, E., Jospin, M., Allard, B. and Segalat, L. (2006) The SLO-1 BK channel of *Caenorhabditis elegans* is critical for muscle function and is involved in dystrophin-dependent muscle dystrophy. *J. Mol. Biol.* **358,** 387-395.

Castillo, P. and Vovlas, N. (2007) *Pratylenchus (Nematoda: Pratylenchidae): Diagnosis, biology, pathogenecity and management*. Leiden, Netherlands: Brill, E. J.

Chitwood, D. J. (2002) Phytochemical based strategies for nematode control. *Annu. Rev. Phytopathol.* **40,** 221-249.

de Boer, J.M., Yan, Y., Smant, G., Davis. E.L., Baum, T.J.,(1998) *In-situ* hybridization to messenger RNA in *Heterodera glycines*. *Journal of Nematol* **30**:309-312.

de Souza Junior, J.D.A., Grossi-de-Sá, M., Rocha, T.L., da Rocha Fragoso, R., de Deus Barbosa, A.E.A., de Oliveira, G.R., Nakasu, E.Y.T., de Sousa, B.A., Pires, N.F., de Alencar Dusi, D.M., Carneiro, R.M.D.G., Romano, E., de Almeida-Engler, J., Engler, G., de Sá, C.M., de Sá, M.F.G. (2011) Ectopic expression of a *Meloidogyne incognita* dorsal gland protein in tobacco accelerates the formation of the nematode feeding site. *Plant Sci* **180**, 276-282.

Doyle, E. A. and Lambert, K. N. (2002) Cloning and characterization of an esophageal-gland-specific pectate lyase from the root-knot nematode *Meloidogyne javanica*. *Mol. Plant-Microbe Interact* **15,** 549-556.

Dubreuil, G., Magliano, M., Deleury, E., Abad, P., Rosso, M.N. (2007) Transcriptome analysis of root knot nematode functions induced in the early stages of parasitism. *New Phytol* **176**:426–436.

Eves-van den Akker, S., Lilley, C. J., Danchin, E. G., Rancurel, C., Cock, P. J., Urwin, P. E.,

and Jones, J. T. (2014). The transcriptome of *Nacobbus aberrans* reveals insights into the evolution of sedentary endoparasitism in plant-parasitic nematodes. *Genome Biol. Evol.* **6**, 2181-2194.

Forbes, J. G., Flaherty, D. B., Ma, K., Qadota, H., Benian, G. M. and Wang, K. (2010) Extensive and modular intrinsically disordered segments in *C. elegans* TTN-1 and implications in filament binding, elasticity and oblique striation. *J. Mol. Biol.* **398,** 672-689.

Fosu-Nyarko, J., and Jones, M. G. K. (2015). Chapter Fourteen -Application of Biotechnology for Nematode Control in Crop Plants. In *Plant Nematode Interactions: A View on Compatible Interrelationships,* (Escobar, C. & Fenoll, C., eds.), Advances in Botanical Research, Academic Press, 2015, **73**, pp. 339-376, ISSN 0065-2296, ISBN 9780124171619,

Frooninckx, L., Van Rompay L., Temmerman L., Van Sinay E., Beets I., Janssen T., Husson, S. J. Husson and Schoofs, L (2012) Neuropeptide GPCRs in *C. elegans*. *Front. Endocrinol.* **3:,** 167.

Fujiwara, M., Ishihara, T. and Katsura, I. (1996) Ligand-gated chloride channel is necessary for correct thermotaxis. *Worm Breeder's Gazette* **14,** 69.

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

García-Añoveros, J., García, J. A., Liu, J.-D. and Corey, D. P. (1998) The nematode degenerin UNC-105 forms ion channels that are activated by degeneration- or hypercontraction-causing mutations. *Neuron* **20,** 1231-1241.

Gladyshev, E. A., Meselson, M., and Arkhipova, I. R. (2008). Massive horizontal gene transfer in bdelloid rotifers. *Science*, **320**, 1210-1213.

Gray, J. M., Karow, D. S., Lu, H., Chang, A. J., Chang, J. S., Ellis, R. E., Marletta, M. A. and Bargmann, C. I. (2004) Oxygen sensation and social feeding mediated by a *C. elegans* guanylate cyclase homologue. *Nature* **430,** 317-322.

Haegeman, A., Joseph, S. and Gheysen, G. (2011) Analysis of the transcriptome of the root lesion nematode *Pratylenchus coffeae* generated by 454 sequencing technology. *Mol. Biochem. Parasit.* **178,** 7-14.

Hajdu-Cronin, Y., Chen, W., Patikoglou, G., Koelle, M. and Sternberg, P. (1999) Antagonism between G(o)alpha and G(q)alpha in *Caenorhabditis elegans*: the RGS protein EAT-16 is necessary for G(o)alpha signaling and regulates G(q)alpha activity. *Genes Dev.* **13,** 1780 - 1793.

Hall, T. A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser* **41,** 95-98.

Hamamouch, N., Li, C., Hewezi, T., Baum, J. T., 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., Howe, P. J., Maier, T. R., Hussey, R. S., Mitchum, M. G., Davis, E. L., & Baum, T. J. (2010). Arabidopsis spermidine synthase is targeted by an effector protein of the cyst nematode *Heterodera schachtii*. *Plant Physiol*, **152**, 968-984.

Hewitson, J.P, Marcus, Y.M., Curwen, R.S., Dowle, A. A., Atmadja, A.K., Ashton, P. D., Wilson, A., Maizels, R. M. (2008) The secretome of the filarial parasite, *Brugia malayi*: proteomic profile of adult excretory-secretory products. *Mol Biochem Parasitol* **160**: 8–21.

Hitoshi, I., Ito, H., Satterlee, J., Sengupta, P., Matsumoto, K. and Mori, I. (2006) Identification of guanylyl cyclases that function in thermosensory neurons of *Caenorhabditis elegans*. *Genetics* **172,** 2239-2252.

Hoppe, P. and Waterston, R. (2000) A region of the myosin rod important for interaction with paramyosin in *Caenorhabditis elegans* striated muscle. *Genetics* **156,** 631 - 643.

Hukema, R. K., Rademakers, S., Dekkers, M. P. J., Burghoom, J. and Jansen, G. (2006) Antagonistic sensory cues generate gustatory plasticity in *Caenorhabditis elegans*. *EMBO J.* **25,** 312-322.

Hunter, S., Rolf, A., Attwood, T. K., Bairoc, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L., Finn, R. D., Gough, J., Haft, D., Nicolas, H., Kahn, D., Kelly, E., Laugard, A., Ivica, L., Lonsdale, D., Lopez, R., Madera, M., Maslen, J., McAnulla, C., McDowall, J., Mistry, J., Mitchell, A., Mulder, N., Natale, D., Orengo, C., Quinn, A. F., Selengut, J. D., Sigrist, C. J. A., Thimma, M., Thomas, P. D., Franck, V., Wilson, D., Wu, C. H. and Yeats, C. (2009) InterPro: the integrative protein signature database. *Nucleic Acids Res.* **37** (Database issue).

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

Jacob, J., Vanholme, B., Haegeman, A. and Gheysen, G. (2007) Four transthyretin-like genes of the migratory plant-parasitic nematode *Radopholus similis*: members of an extensive nematode-specific family. *Gene* **402,** 9-19.

Jaouannet, M., Magliano, M., Arguel, M. J., Gourgues, M., Evangelisti, E., Abad, P. and Rosso, M. N. (2012) The root-knot nematode calreticulin Mi-CRT is a key effector in plant defense suppression. *Mol.Plant-Microbe Interact.* **26,** 97-105.

Jaubert, S., Laffaire, J.B., Abad, P., Rosso, M.N., (2002a) A polygalacturonase of animal origin isolated from the root-knot nematode *Meloidogyne incognita*. FEBS Lett. **522**, 109-112.

Jaubert, S., Ledger, T.N., Laffaire, J.B., Piotte, C., Abad, P., Rosso, M.N. (2002b) Direct identification of stylet secreted proteins from root-knot nematodes by a proteomic approach. *Mol Biochem Parasitol* **121**:205–211.

Jaubert, S., Laffaire, J. B., Ledger, T. N., Escoubas, P., Amri, E. Z., Abad, P., and Rosso, M. N. (2004). Comparative analysis of two 14-3-3 homologues and their expression pattern in the root-knot nematode *Meloidogyne incognita*. *Int. J. Parasitol.* **34**, 873-880.

Jaubert, S., Milac, A. L., Petrescu, A. J., de Almeida-Engler, J., Abad, P., and Rosso, M. N. (2005). *In planta* secretion of a calreticulin by migratory and sedentary stages of root-knot nematode. *Mol. Plant Microbe Interact.*,**18**, 1277-1284.

Jones, J.T., Reavy, B., Smant. G., Prior, A.E. (2004) Glutathione peroxidases of the potato

cyst nematode *Globodera rostochiensis. Gene* **324**:47–54.

Jones, M. G. K. (1981) Host cell responses to endoparasitic nematode attack: structure and function of giant cells and syncytia. *Ann. Appl. Biol.* **97,** 353-372.

Jones, M. G. K. and Fosu-Nyarko, J. (2014) Molecular biology of root lesion nematodes (*Pratylenchus* spp.) and their interaction with host plants. *Ann. Appl. Biol.* **164,** 163-181.

Jones, M. G. K. and Fosu-Nyarko, J. (2014) Molecular biology of root lesion nematodes (*Pratylenchus* spp.) and their interaction with host plants. *Ann. Appl. Biol.* **164,** 163-181.

Jones, M G K and Goto, D B, (2011). 'Root-Knot Nematodes and Giant Cells', Chapter 5 in 'Genomics and Molecular Genetics of Plant Nematode Interactions', Eds Jones, J., Gheysen, G. and Fenoll, C., Springer; pp 83-100.

Jones, M. G. K. and Northcote, D. H. (1972) Nematode-induced syncytium-a multinucleate transfer cell. *J. Cell. Sci.* **10,** 789-809.

Kanazawa, T., Zappaterra, M. D., Hasegawa, A., Wright, A. P., Newman-Smith, E., Buttle, K. F., McDonald, K., Mannella, C. A. and van der Bliek, A. M. (2008) The *C. elegans* Opa1 homologue EAT-3 is essential for resistance to free radicals. *PLoS Genet.* **4**.

Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M. and Hirakawa, M. (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* **38 (**Database issue), D355-360.

Khan, S. A. (1959) Pathogenic effects of *Pratylenchus zeae* on sugarcane. *Phytopathology* **59,** 543.

Kimata, T., Tanizawa, Y., Can, Y., Ikeda, S., Kuhara, A. and Mori, I. (2012) Synaptic polarity depends on phosphatidylinositol signaling regulated by myo-inositol monophosphatase in *Caenorhabditis elegans*. *Genetics* **191,** 509-521.

Koelle, M. and Horvitz, H. (1996) EGL-10 regulates G protein signaling in the *C. elegans* nervous system and shares a conserved domain with many mammalian proteins. *Cell* **84,** 115 - 125.

Komatsu, H., Mori, I., Rhee, J.-S., Akaike, N. and Ohshima, Y. (1996) Mutations in a cyclic nucleotide gated channel lead to abnormal thermosensation and chemosensation in *C. elegans*. *Neuron* **17,** 707-718.

Kumar, M., Gantasala, N. P., Roychowdhury, T., Thakur, P. K., Banakar, P., Shukla, R. N., Jones, M. G. K. and Rao, U. (2014). *De novo* transcriptome sequencing and analysis of the cereal cyst nematode, *Heterodera avenae*. PLoS ONE **9**: e96311. doi:10.1371/journal.pone.0096311.

Lackner, M., Nurrish, S. and Kaplan, J. (1999) Facilitation of synaptic transmission by EGL-30 Gq alpha and EGL-8 PLC beta: DAG binding to UNC-13 is required to stimulate acetylcholine release. *Neuron* **24,** 335 - 346.

Lee, C., Chronis, D., Kenning, C., Peret, B., Hewezi, T., Davis, E. L., Baum, T.J., Hussey, R., Bennet, M and Mitchum, M. G. (2011). The novel cyst nematode effector protein 19C07 interacts with the Arabidopsis auxin influx transporter LAX3 to control feeding site development. *Plant Physiol*, **155**, 866-880.

Lee, R. Y. N., Sawin, E. R., Chalfie, M., Horvitz, H. R. and Avery, L. (1999) EAT-4, a

homolog of a mammalian sodium-dependent inorganic phosphate cotransporter, is necessary for glutamatergic neurotransmission in *Caenorhabditis elegans*. *J. Neurosci.* **19,** 159-167.

L'Etoile, N. D. and Bargmann, C. I. (2000) Olfaction and odor discrimination are mediated by the *C. elegans* guanylyl cyclase ODR-1. *Neuron* **25,** 575-586.

Li, Z., Liu, X., Chu, Y., Wang, Y., Zhang, Q., Zhou, X, (2011) Cloning and characterization of a 2-Cys peroxiredoxin in the pine wood nematode, *Bursaphelenchus xylophilus*, a putative genetic factor facilitating the infestation. *Int. J. Biol. Sci.* **7**:823.

Lilley, C.J., Goodchild, S.A., Atkinson, H.J., Urwin, P.E. (2005) Cloning and characterisation of a *Heterodera glycines* aminopeptidase cDNA. *Int J Parasitol* **35**:1577–1585.

Lozano-Torres, J. L., Wilbers, R. H. P., Gawronski, P., Boshoven, J. C., Finkers-Tomczak, A., Cordewener, J. H. G., America, A. H. P., Overmars, H. A., Van 't Klooster, J. W., Baranowski, L., Sobczak, M., Ilyas, M., van der Hoorn, R. A. L., Schots, A., de Wit, P. J. G. M., Bakker, J., Goverse, A. and Smant, G. (2012) Dual disease resistance mediated by the immune receptor Cf-2 in tomato requires a common virulence target of a fungus and a nematode. *PNAS* **109,** 10119-10124.

Lu, W., Egerton, G. L., Bianco, A. E., and Williams, S. A. (1998) Thioredoxin peroxidase from *Onchocerca volvulus*: a major hydrogen peroxide detoxifying enzyme in filarial parasites. *Mol. Biochem. Parasitol.* **91**, 221-235.

Lusk, R.W. (2014). Diverse and widespread contamination evident in the unmapped depths of high throughput sequencing data. PLOS One 9: e11808. Doi10.1371/journal.pone.0110808

Manion, M., LeVan, K., Ni, S., Wu, Y. J. and Liu, C. J. (2009) Reducing Error in Next Generation Sequencing Data with NextGENe Software's Condensation Tool.

Marroquin, L. D., Elyassnia, D., Griffitts, J. S., Feitelson, J., S and Aroian, R. V. (2000) B*acillus thuringiensis* (Bt) toxin susceptibility and isolation of resistance mutants in the nematode *Caenorhabditis elegans Genetics* **155,** 1693-1699.

Matsubayashi, Y. (2011). Post-translational modifications in secreted peptide hormones in plants. *Plant Cell Physiol*, **52**, 5-13.

Maule, A. G. and Curtis, R. (2011) Parallels between plant and animal parasitic nematodes. In: *Genomics and Molecular Genetics of Plant-Nematode Interactions.* (Jones, J. T., Gheysen, G. and Fenoll, C., eds.) Springer, pp. 234.

Moreno, Y., Geary, T.G. (2008) Stage- and gender-specific proteomic analysis of *Brugia malayi* excretory-secretory products. *PLoS Negl Trop Dis* **2**: e326.

Nathalie, P., Zugasti, O., Wong, D., Couillault, C., Leopold-Kurz, C., Schulenburg, H. and Ewbank, J. J. (2008) Anti-fungal innate immunity in *C. elegans* is enhanced by evolutionary diversification of antimicrobial peptides. *PLoS Pathog* **4**.

Nelson, L., Rosoff, M. and Li, C. (1998) Disruption of a neuropeptide gene, flp-1, causes multiple behavioral defects in *Caenorhabditis elegans*. *Science* **281,** 1686 - 1690.

Neveu, C., Abad, P., Castagnone-Sereno, P. (2003) Molecular cloning and characterization of an intestinal cathepsin L protease from the plant-parasitic nematode *Meloidogyne incognita*, *Physiol. Mol. Plant Pathol*, **63**: 159-165.

Nicol, P., Gill, R., Fosu-Nyarko, J. and Jones, M. G. K. (2012) *de novo* analysis and functional classification of the transcriptome of the root lesion nematode, *Pratylenchus thornei,* after 454 GS FLX sequencing. *Int. J. Parasitol.* **42,** 225-237.

Parra, G., Bradnam, K., Korf, I. (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23,** 1061-1067.

Patel, N., Hamamouch, N., Li, C.Y., Hewezi, T., Hussey, R.S., Baum, T., Mitchum, M., Davis, E.L. (2010) A nematode effector protein similar to annexins in host plants. *J. Exp. Bot.* **61**:235–248

Perry, R. N. (1996) Chemoreception in plant parasitic nematodes. *Annu. Rev. Phytopathol.* **34,** 181-199.

Plenefisch, J. D., Zhu, X. and Hedgecock, E. M. (2000) Fragile skeletal muscle attachments in dystrophic mutants of *Caenorhabditis elegans*: isolation and characterization of the mua genes. *Development* **127,** 1197-1207.

Prior, A., Jones, J.T., Blok, V.C., Beauchamp, J., McDermott, L., Cooper, A., Kennedy, M.W. (2001) A surface-associated retinol- and fatty acid-binding protein (Gp-FAR-1) from the potato cyst nematode *Globodera pallida*: lipid binding activities, structural analysis and expression pattern. *Biochem J* **356**:387–394

Reinitz, C. A., Herfel, H. G., Messinger, L. A. and Stretton, A. O. W. (2000) Changes in locomotory behavior and cAMP produced in *Ascaris suum* by neuropeptides from *Ascaris suum* or *Caenorhabditis elegans*. *Mol. Biochem. Parasit.* **111,** 185-197.

Reynolds, A. M., Dutta, T. K., Curtis, R. H. C., Powers, S. J., Gaur, H. S. and Kerry, B. R. (2011) Chemotaxis can take plant-parasitic nematodes to the source of a chemo-attractant via the shortest possible routes. *J. Royal Society Interface* **8,** 568-577.

Roayaie, K., Crump, J. G., Sagasti, A. and Bargmann, C. I. (1998) The Gα protein ODR-3 mediates olfactory and nociceptive function and controls cilium morphogenesis in *C. elegans* olfactory neurons. *Neuron* **20,** 55-67.

Robatzek, M., Niacaris, T., Steger, K., Avery, L. and Thomas, J. (2001) *eat-11* encodes GPB-2, a G beta (5) ortholog that interacts with G(o) alpha and G(q) alpha to regulate *C. elegans* behavior. *Curr. Biol.* **11,** 288 - 293.

Robertson, L., Robertson, W. M., Sobczak, M., Helder, J., Tetaud, E., Ariyanayagam, M. R., Ferguson, M. A. J, Fairlamb, A., Jones, J. T. (2000) Cloning, expression and functional characterisation of a peroxiredoxin from the potato cyst nematode *Globodera rostochiensis*. *Mol Biochem Parasitol* **111**:41–49.

Sacco, M.A., Koropacka, K., Grenier, E., Jaubert, M. J., Blanchard, A., Goverse, A., Smant, G., Moffett, P. (2009) The cyst nematode SPRYSEC protein RBP-1 elicits Gpa2- and RanGAP2-dependent plant cell death. *PLoS Pathogens* 5:e1000564.

Satterlee, J. S., Ryu, W. S. and Sengupta, P. (2004) The CMK-1 CaMKI and the TAX-4 Cyclic nucleotide-gated channel regulate thermosensory neuron gene expression and function in *C. elegans*. *Curr. Biol.* **14,** 62-68.

Savage, C., Hamelin, M., Culotti, J. G., Coulson, A., Albertson, D. G. and Chalfie, M. (1989) *mec-7* is a β-tubulin gene required for the production of 15-protofilament microtubules in *Caenorhabditis elegans*. *Genes Dev.* **3,** 870-887.

Semblat, J.-P., Rosso, M.-N., Hussey, R. S., Abad, P. and Castagnone-Sereno, P. (2001) Molecular cloning of a cDNA encoding an amphid-secreted putative avirulence protein from the root-knot nematode *Meloidogyne incognita*. *Mol. Plant-Microbe Interact.* **14,** 72-79.

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

Soumi, J., Gheysen, G. and Subramaniam, K. (2012) RNA interference in *Pratylenchus coffeae*: knockdown of Pc-pat-10 and Pc-unc-87 impedes migration. *Mol. Biochem. Parasit.* **186,** 51-59.

Starich, T. A., Lee, R. Y., Panzarella, C., Avery, L. and Shaw, J. E. (1996) e*at-5* and *unc-7* represent a multigene family in *Caenorhabditis elegans* involved in cell-cell coupling. *J. Cell Biol.* **134,** 537-548.

Stirling, G. R. (1991) *Biological control of plant parasitic nematodes*. Brisbane, Australia: CAB International.

Tan, J.-A. C. H., Jones, M. G. K. and Fosu-Nyarko, J. (2013) Gene silencing in root lesion nematodes (*Pratylenchus* spp.) significantly reduces reproduction in a plant host. *Exp. Parasitol.* **133,** 166-178.

Terami, H., Williams, B. D., Kitamura, S.-i., Sakube, Y., Matsumoto, S., Doi, S., Obinata, T. and Kagawa, H. (1999) Genomic organization, expression, and analysis of the troponin C gene *pat-10* of *Caenorhabditis elegans*. *J. Cell Biol.* **146,** 193-202.

Trevathan, L. E., Cuarezma-Teran, J. A. and Gourley, L. M. (1985) Relationship of plant-nematodes and edaphic factors in colombian grain sorghum production. *Nematropica* **15,** 145-153.

Vowels, J. J. and Thomas, J. H. (1994) Multiple chemosensory defects in *daf-11* and *daf-21* mutants of *Caenorhabditis elegans*. *Genetics* **138,** 303-316.

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*. *Mol. Plant-Microbe Interact.* **14,** 536-544.

Zacheo, G. and Bleve-Zacheo, T. (1988) Involvement of superoxide dismutases and superoxide radicals in the susceptibility and resistance of tomato plants to *Meloidogyne incognita* attack. *Physiol. Mol. Plant Pathol*. **32**:313-22

Zacheo, G., Bleve-Zacheo, T., Pricolo, G. (1987) Metabolic changes in enzyme levels in potato roots infested by potato cyst-nematodes, *Globodera pallida* (Pa3) and *Globodera rostochiensis* (Rol). *Nematol. Medit*. **15** :293-302

Zunke, U. (1990) Observation on the invasion and endoparasitic behaviour of the root lesion nematodes *Pratylenchus penetrans*. *J. Nematol.* **22,** 309-322.

Supporting information

Supplementary Methods, S1: Details of methodology for functional characterisations of *P. zeae* transctipts putatively encoding five proteins with movement and parasitism functions.

Table S1: Number of *P. zeae* transcripts putatively encoding known CAZyme families in the CAZy database.

Table S2: Transcripts of *P. zeae, P. thornei, H. avenae* and genomic sequences of *M. incognita, M. hapla* and *G. pallida* similar to the EuKaryotic clusters of Orthologous Groups (KOGs) of genes.

Table S3:

Table S3a: TBLASTX matches to *P. zeae* transcript putatively encoding a secreted 1,4-beta endoglucanase, *Pzeng*.

Table S3b: TBLASTX matches to *P. zeae* transcript putatively encoding troponin c, *Pzpat-10*.

Table S3c: TBLASTX matches to *P. zeae* transcript putatively encoding calponin, *Pzunc-87*

Table S3d: TBLASTX matches to *P. zeae* transcript putatively encoding a secreted calreticulun, *Pzcrt-1.*

Table S3e: TBLASTX matches to *P. zeae* transcript putatively encoding a secreted SPRY containing protein, *Pzsxp.*

Legends to Tables and Figures

Table 1: *P. zeae* transcripts putatively encoding CAZymes similar to those of plant parasitic nematodes.

Table 2*: P. zeae* transcripts with best matches to ten common genes for nematodes with different lifestyles.

Table 3: *P. zeae* transcripts similar to putative parasitism genes of PPNs and *Ascaris suum*.

Table 4: *P. zeae* transcripts with putative homology to proteins/peptides of the secretomes of *M. incognita, H. glycines* and the secretory-excretory products of *B. malayi*.

Fig. 1: Sequence characteristics of the *P. zeae* transcriptome: (A) Distribution of contig lengths (B) Distribution of read lengths.

Fig. 2: Functional classification of *P. zeae* transcripts. A) Percentage of transcripts mapping to GO terms with molecular functions B) Percentage of transcripts matching

terms for biological processes C) Percentage of transcripts matching terms with cellular components D) Classification of transcripts into biochemical pathways using KEGG.

Fig. 3: *P. zeae* transcripts matching genes/ESTs of 16 nematode species.

Fig. 4: Distribution of *P. zeae* transcripts among FLNs, APNs, PPNs and *P. pacificus*. (A) Number of *P. zeae* contigs and (B) singletons with matches to genes/ ESTs of the FLNs, APNs and PPNs. (C and D): Number of *P. zeae* contigs (C) and singletons (D) with similarities to *P. pacificus* genes. FLNs: *C. elegans, C. briggssae, C. remanei, APN: A. suum, B. malayi, PPN: H. glycines, H. schachtii, G. pallida, G. rostochiensis, M. incognita, M. hapla, M. javanica, R. similis, B. xylophilus, D. africanus.*

Fig. 5: Number of genes/ESTs with unique identifiers of 16 different nematodes matching transcripts of *P. zeae* (total bit score >100).

Fig. 6: Comparative analysis of *P. zeae* transcripts with transcriptomes and genomes of six PPNs. (A) TBLASTX matches of *P. zeae* transcripts to TSA of *P. thornei*, transcriptome of *H. avenae* and genomic contigs of *H. glycines, G. pallida, M. incognita,* and *M. hapla.* (B) Comparison of homologues of CEGs in the transcriptomes of *P. zeae*, *P. thornei*, *H. avenae* and genomes of *H. glycines, G. pallida, M. incognita,* and *M. hapla.*

Fig. 7: Number of *P. zeae* transcripts similar to *C. elegans* genes with roles in structural integrity of the cuticle and movement, and those involved in responses to stimuli.

Fig. 8: Functional characterisation of *Pzeng*, *Pzcrt-1* and *Pzsxp, Pzpat-10* and *Pzunc-87.* (I) *In situ* hybridisation of *Pzeng* (A), *Pzcrt-1* (B), *Pzsxp* (C), sense probe of *Pzeng* (D), *Pzpat-10* (E) and *Pzunc-87* (F). Scale bar represents 50 μm. (II) Fold reduction of gene expression of *Pzeng*, *Pzcrt-1* and *Pzsxp* in nematodes 16 h after incubation with corresponding dsRNA compared to expression in nematodes soaked with dsGFP. (III) Reduction in number of *P. zeae* on carrot discs 5 weeks after soaking with dsPzeng, dsPzcrt-1 and dsPzsxp. Note: bars with the same letters represent significantly different means ($p<0.05$).

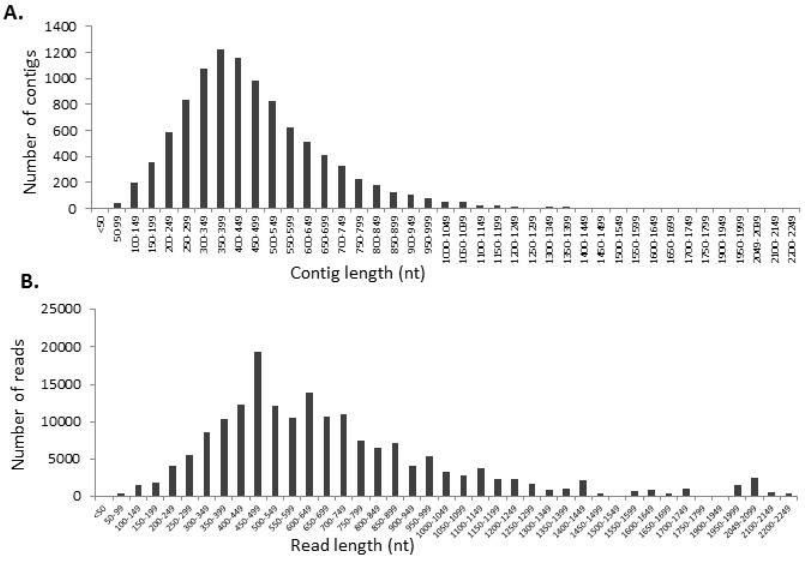

Fig. 1: Sequence characteristics of the *P. zeae* transcriptome: (A) Distribution of contig lengths (B) Distribution of read lengths.

MPP_12301_F1

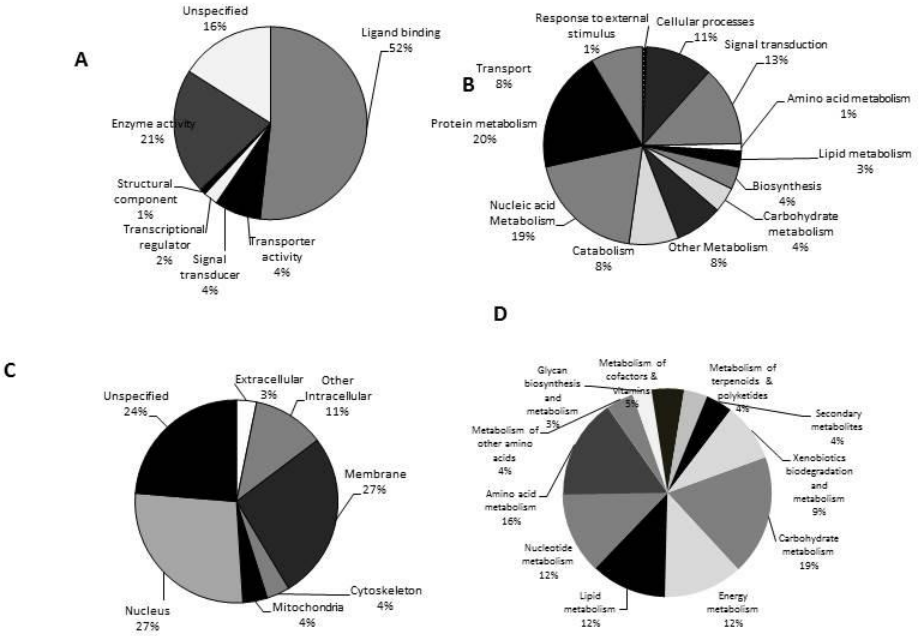

Fig. 2: Functional classification of *P. zeae* transcripts. A). Transcripts mapping to GO terms with molecular functions, B) Transcripts terms for biological processes. C). Transcripts terms for cellular components. D). Classification of transcripts into biochemical pathways using KEGG.

MPP_12301_F2

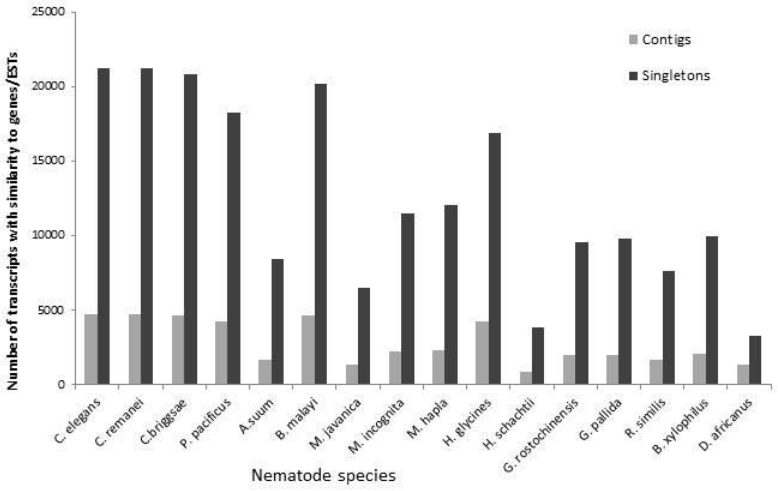

Fig. 3: *P. zeae* transcripts matching genes/ESTs of 16 nematodes species

MPP_12301_F3

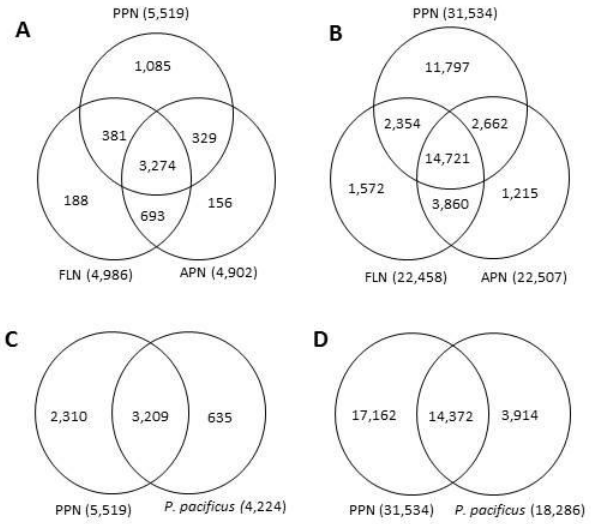

Fig. 4: Distribution of *P. zeae* transcripts among FLNs, APNs, PPNs and *P. pacificus*.
A). Number of *P. zeae* contigs and (B) singletons with matches to genes/ESTs of the FLNs, APNs and PPNs. (C and D): Number of *P. zeae* contigs (C) and singletons (D) with similarities to *P. pacificus* genes. FLNs: *C. elegans, C. briggssae, C. remanei*; APNs: A. *suum, B. malayi*; PPN: *H. glycines, H. schachtii, G. pallida, G. rostochinensis, M. incognita, M. hapla, M. javanica, R. similis, B. xylophilus, D. africanus*.

MPP_12301_F4

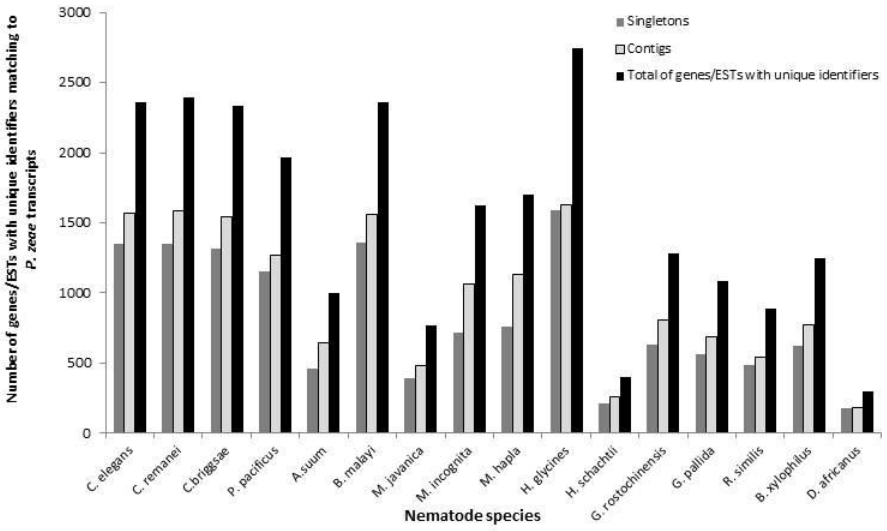

Fig. 5: Number of genes/ESTs with unique identifiers of 16 different nematodes matching transcripts of *P. zeae* (total bit score >100).

MPP_12301_F5

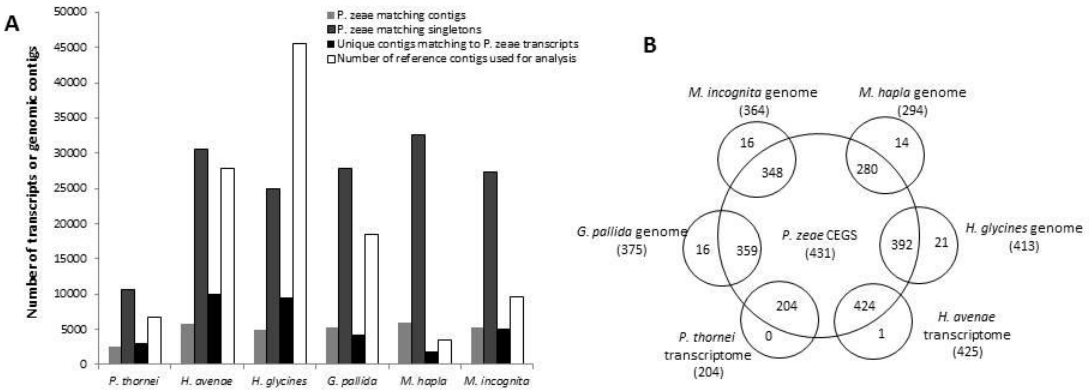

Fig. 6. Comparative analysis of *P. zeae* transcripts with transcriptomes of and genomes of six PPNs.
A). TBLASTX matches of *P. zeae* transcripts to TSA of *P. thornei*, transcriptome of *H. avenae* and genomic contigs of *H. glycines*, *G. pallida*, *M. incognita* and *M. hapla.* B). Comparison of homologues of CEGs in the transcriptomes of *P. zeae*, *P. thornei* and *H. avenae* and genomes of *H. glycines*, *G. pallida*, *M. incognita* and *M. hapla.*

MPP_12301_F6

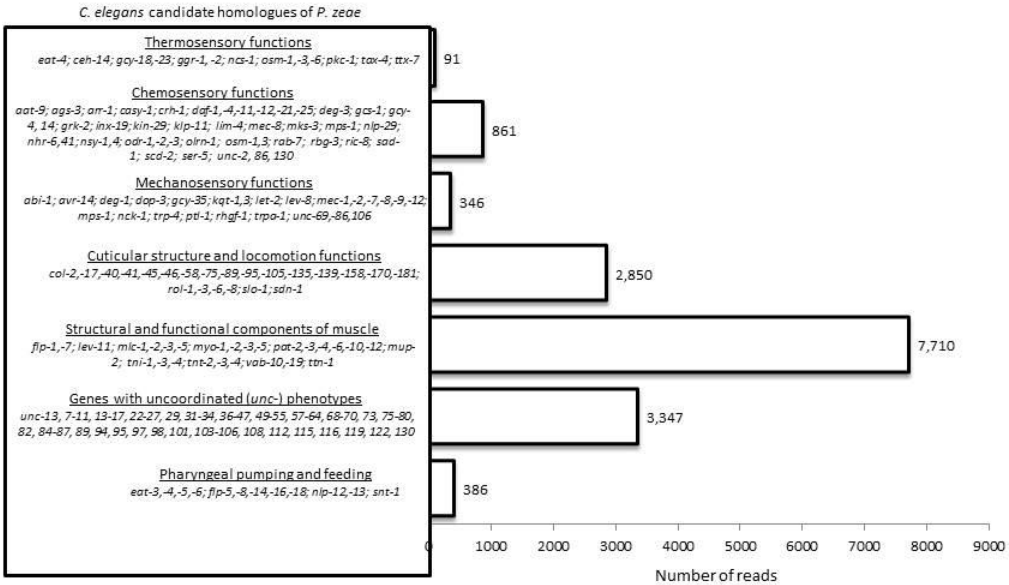

Fig. 7: Number of *P. zeae* transcripts similar to *C. elegans* genes with roles in structural integrity of the cuticle and movement, and those involved in responses to stimuli.

MPP_12301_F7

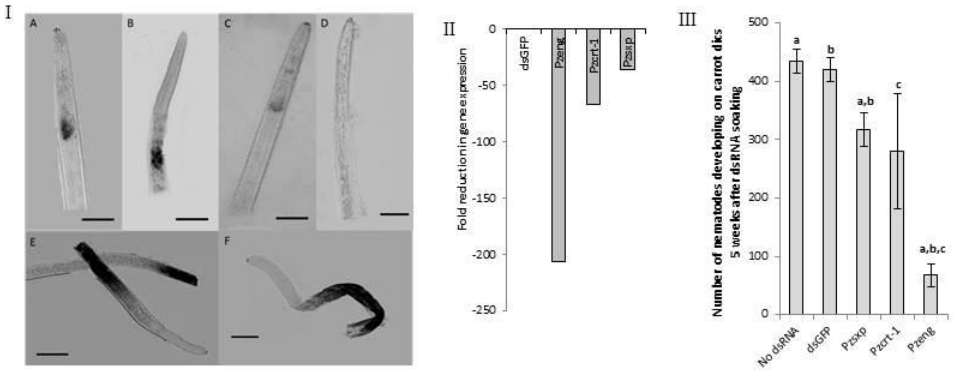

Fig. 8: Functional characterisation of *Pz-eng, Pz-crt, Pz-SXP-RAL2, Pz-pat-10* and *Pz-unc-87*.
I) *In situ* hybridisation of *Pzeng* (A), *Pzcrt-1* (B), *Pzsxp* (C), sense probe of *Pzeng* (D), *Pzpat-10* (E) and *Pzunc-87* (F). Scale bar represents 50µm.
II) Fold reduction of gene expression of *Pzeng, Pzcrt-1* and *Pzsxp* in nematodes 16 h after incubation with corresponding dsRNA compared to expression in those soaked with dsGFP.
III) Reduction in number of developing *P. zeae* on carrot discs 5 weeks after soaking with dsPzeng, dsPzcrt-1 and dsPzsxp. Note: bars with same letters represent significantly different means ($p<0.05$).

MPP_12301_F8

Table 1: *P. zeae* transcripts putatively encoding CAZymes similar to those of plant parasitic nematodes

CAZy family	CAZymes putatively encoded by *P. zeae* transcripts	Number of matching transcripts	BLASTX characteristics				
			Organism with best CAZyme match	NCBI Accession number of CAZyme with greatest BLASTX % identity	Lowest E-value	Greatest BLASTX hit identity (%)	Greatest positive (%) to best matching accession number
GH5	Cellulase, partial	1	*Aphelenchoides fragariae*	AFI63769	4.89E-18	80.56	94.44
GH5	Beta-1,4-endoglucanase	2	*Pratylenchus penetrans*	BAB68523	1.76E-06	69.57	86.96
GH5	Beta-1,4-endoglucanase, partial	5	*Pratylenchus pratensis*	AER27775	1.26E-11	83.33	95.83
GH5	Beta-1,4-endoglucanase, partial	7	*Pratylenchus vulnus*	AER27785	1.85E-07	94.12	100
GH5	GHF5 beta-1,4-endoglucanase	3	*Rotylenchulus reniformis*	ADM72857	1.59E-08	90	100
GH5, CBM2	Beta-1,4-endoglucanase	15	*Ditylenchus destructor*	ADW77528	1.05E-17	53.85	72.31
GH5, CBM2	GHF5 endoglucanase precursor	9	*Pratylenchus coffeae*	ABX79356	7.35E-24	95.35	96.3
GH5, CBM2	GHF5 endo-1,4-beta-glucanase precursor	3	*Radopholus similis*	ABV54446	6.73E-11	50	75
PL3	Pectate lyase	2	*Heterodera glycines, Heterodera schachtii*	ADW77534; ABN14273	8.54E-08	69.77	79.07
GT20	Putative trehalose 6-phosphate synthase	15	*Aphelenchus avenae*	CAH18869; CAH18871; CAH18873; CAH18874	1.95E-73	96.29	94.73
GT2	Chitin synthase	1	*Meloidogyne artiellia*	AAG40111	1.09E-10	83.33	84

Table 2*: P. zeae* transcripts with best matches to ten common genes of nematodes with different lifestyles.

Gene description	Wormbase Gene Id	Number of matching contigs, singletons	Length of longest contig (nt)	Total bit scores of longest contig to genes of nematode species				
				C. elegans	***B. malayi***	***P. pacificus***	***M. incognita***	***H. glycines***
Heat shock hsp70 proteins	WBGene00009692	2, 3	1301	994	895	771	1041	932
Actin related protein of the conserved Arp2/3 complex (arx-4)	WBGene00021170	1, 1	925	1221	414	972	1168	1059
Protein kinase N (pkn-1)	WBGene00009793	1, 10	1055	1330	438	228	1068	804
Succinate dehydrogenase (ubiquinone) iron sulphur protein (dhb-1)	WBGene00006433	1, 3	926	1294	403	692	1586	1593
Guanine nucleotide-binding protein (goa-1)	WBGene00001648	3, 13	921	1582	335	1299	1843	1076
Actin (act-2)	WBGene00000064	2, 16	864	1664	1079	1472	1522	1848
Casein Kinase I (kin-19)	WBGene00002202	4, 14	1383	1491	498	903	1197	1414
ATP-dependent DEAD-box RNA helicase (cgh-1)	WBGene00000479	2, 9	771	1378	447	1247	205	1114
Casein kinase (kin-3)	WBGene00002191	1, 4	1224	1153	1202	1279	1478	798
Chemokine (C-C motif) receptor 4 (ccr-4)	WBGene00000376	3, 5	1264	1793	592	154	770	1188

Table 2: *P. zeae* transcripts similar to putative parasitism genes of PPNs and *Ascaris suum*.

Nematode parasitism gene	Number of matching transcripts (contigs, singletons)	Length of best matching transcript (nt)	Nematode species with matches to *P. zeae* transcripts (*nematode with best match to most similar transcript)											TBLASTX results for best matching transcript and accession number			Example of evidence of confirmed or suggested function/role in parasitism
			Mi	Mj	Mh	Gp	Gr	Hg	Hs	Rs	Da	As	Pv	Total Score	Query coverage (%)	Accession number of best matching gene (length, nt) *	
Sec-2 protein	3, 26	923	+	+	+	+*	+	+	+	+	+	-	-	1834	77	BM415209 (1338)	Prior et al., 2001
RANBP domains	1, 9	443	-	-	-	-	-	+*	-	-	-	-	-	385	84	CA939458.1 (484)	Sacco et al., 2009
Superoxide dismutase	2, 18	624	+	-	+*	-	+	+	+	+	+	+	-	930	71	CF803336 (469)	Bellafiore et al., 2008
Thioredoxin	7, 58	758	-	+	-	-	+*	-	+	-	-	+	-	873	59	AW505887 (537)	Lu et al., 1998
Ubiquitin extension protein	2, 3	520	-	-	-	-	+*	-	-	-	-	-	-	965	91	EE269774 (527)	Jones et al., 2009
Transthyretin-like protein (TTL)	14, 107	455	+	+	-	+	+	-	-	+*	-	-	-	1230	88	EY190294.1 (488)	Jacob et al., 2007
S-phase kinase-associated protein 1 (SKP-1)	4, 0	459	+*	+	+	-	-	-	-	-	-	-	-	365	79	CF980621.1 (652)	Gao et al., 2003
14-3-3b protein	5, 27	618	+	+	+	-	+*	+	+	+	+	+	-	982	88	EE267463.1 (594)	Jaubert et al., 2002b
Aminopeptidase	9, 48	533	+	-	+	-	+	+	-	+	+	+	-	997	81	CA995734.1 (701)	Lilley et al., 2005
Annexin	1, 30	581	+	-	+*	+	+*	+	-	-	+	+	-	1415	96	BM345703.1 (583)	Jones et al., 2009
^Cathepsins B, L, S, D, and Z	12, 94	669	+	+	+	+*	+	+	+	+	+	+	-	1653	96	BM415989.1 (1262)	Neveu et al., 2003
Cellulose binding protein	1, 1	379	+*	-	-	-	-	-	-	-	-	-	-	159	64	AW828050.1 (548)	Adam et al., 2008
Endoglucanases and precursors	3, 45	339	+	+	+	+	+	+	+	+	+	-	+*	916	83	CV200806.1 (507)	Smant et al., 1998
Galectin	11, 57	448	+	+	+	+	+	+	+*	+	+	+	-	1433	98	CD750386.1 (508)	Dubreuil et al., 2007
Glutathione peroxidase	3, 25	775	+*	+	+	+	+	+	-	+	-	+	-	906	50	AW829298.1 (537)	Jones et al., 2004
Glutathione S-transferase	4, 51	448	+	+	+	+	+	+	-	+*	+	+	-	1191	91	EY194552.1 (613)	Dubreuil et al., 2007
Calnexin/calreticulin	0, 11	390	+	-	-	+	-	+*	-	-	+	+	-	517	97	CB380179.1 (554)	Jaouannet et al., 2013
Pectate lyase	0, 3	219	-	-	-	-	+	+*	+	-	-	-	-	249	75	CB279891.1 (499)	Doyle and Lambert, 2003
Peroxiredoxin	0, 14	204	+	-	+	+	+	+*	-	-	-	-	-	516	96	CB299476.1 (612)	Robertson et al., 2000
Polygalacturonase	0, 3	373	+	+	+*	-	-	-	-	-	-	-	-	273	92	BU095396.1 (566)	Jaubert et al., 2002a
Venom allergen-like proteins (Vap-1)	0, 3	245	-	-	-	+	+	-	-	-	+	-	-	599	92	BM416306.1 (1284)	Lozano-Torres et al., 2012

^The number of contigs and singletons represent matches to cathepsins B, L, S, D, and Z from the nematodes indicated in column 4; TBLASTX results is for cathepsin L of *G. pallida*.

+ indicates presence of a similar sequence (ESTs or gene) of the respective nematode in the NCBI database

- indicates absence of a similar sequence (ESTs or gene) of the respective nematode in the NCBI database

Mi= *M. incognita*; Mj=*M. javanica*; Mh=*M. hapla*; Gp=*G. pallida*; Gr=*G. rostochiensis;* Hg=*H. glycines*; Hs=*H. schachtii*; Rs=*R. similis*, Da=*D. africanus:* Pv=*P. vulnus;* As=*Ascaris suum*

Table 4*: P. zeae* transcripts with putative homology to proteins/peptides of the secretomes of *M. incognita, H. glycines* and secretory-excretory products of *B. malayi*.

	Secretome of nematode					
	Secretome of sedentary		Excretory-secretory products of *B. malayi* life			
Proteins	***M. incognita*[1]**	***H. glycines*[2]**	**Host-parasite interface[3]**	**Microfilariae, female and males[4]**	**Adult life stage[5]**	**Number of matching contigs (singletons)**
Structural/cytoskeleton/Nuclear						
Actin	YES	-	-	-	YES	13 (186)
Calmodulin	YES	-	-	-	YES	9 (26)
Tropomyosin family protein	YES	-	-	-	YES	6 (120)
High mobility group (HMG) protein	YES	-	-	-	YES	10 (37)
Intermediate filament protein	-	-	YES	-	-	3 (12)
Cuticle collagen	-	-	YES	-	-	8 (46)
N-Heparan sulphate	-	-	YES	-	-	0 (0)
Troponin C-like protein	YES	-	-	-	-	2 (0)
Myosin regulatory light chain	YES	-	-	-	-	7 (72)
Polygalacturonase	YES	-	-	-	-	0 (1)
Cellulose binding protein	YES	YES	-	-	-	1 (1)
Twitchin (unc-22)	YES	-	YES	-	-	2 (42)
Guanylyl cyclase	-	YES	-	-	-	2 (47)
Chitinase	-	YES	-	-	-	0 (8)
Histone H4	-	-	YES	-	-	1 (14)
Msp-1	-	YES	-	YES	YES	2 (13)
Energy metabolism						
6-phosphofructokinase	-	-	-	-	YES	0 (2)
Enolase	YES	-	-	YES	YES	1 (11)
Fructose-bisphosphate aldolase 1	-	-	-	YES	-	0 (0)
Transaldolase	-	-	YES	-	-	1 (5)
ATP synthase (atp-2)	YES	-	-	-	-	17 (72)
Adenylate cyclase	-	-	YES	-	-	2 (9)
Nucleoside diphosphate kinase	-	-	YES	-	-	3 (6)
Triosephosphate isomerase	YES	-	YES	YES	YES	1 (1)
Guanylate kinase	-	-	YES	-	-	8 (11)
Phosphoglycerate mutase	-	-	-	YES	-	1 (4)
Protein digestion, folding and fate/calcium binding						
Aminopeptidase	-	-	YES	YES	-	3 (20)
Calreticulin	YES	-	-	-	YES	0 (6)
Serine carboxypeptidase	-	-	-	-	YES	2 (10)
Ubiquitin (extension protein)	YES	YES	-	-	YES	0 (0)
Ubiquitin-like protein SMT3	YES	-	-	-	YES	2 (8)
Peptidyl-prolyl cis-trans isomerase	YES	-	YES	-	YES	3 (23)
Serine/threonine protein	-	-	YES	-	YES	9 (72)
Cytochrome c type 1 protein	-	-	YES	-	-	2 (0)
SKP-1	-	YES	-	-	-	3 (22)
Putative trypsin inhibitor	-	-	-	YES	-	0 (13)
Calsequestrin family protein	-	-	-	YES	YES	1 (5)
Cyclophilin	YES	-	-	-	YES	3 (28)
Protein disulphide isomerase	-	-	-	-	YES	4 (10)
Cytosol stress response and antioxidants						
14-3-3b	YES	-	-	-	YES	5 (24)
Hsp 70	YES	-	-	-	YES	3 (9)
Heat shock 90	-	-	-	YES	-	5 (26)
Glutathione peroxidase	YES	-	YES	YES	YES	4 (16)
Thioredoxin	-	-	YES	-	YES	3 (14)
Superoxide dismutase	YES	-	YES	YES	YES	2 (16)
Translationally controlled tumor	YES	-	-	-	YES	2 (1)
Annexin	YES	YES	-	-	-	1 (17)
Glutathione -S-transferase	YES	-	-	-	-	3 (12)
Glutathione reductase	-	-	YES	-	-	3 (0)
Glutathione synthetase	YES	-	-	-	-	3 (6)
	-	-	-	-	-	
Lectins and glycosyltransferases						
Galectin	YES	-	YES	YES	YES	7 (33)

N-acetylglucosaminyltransferase	-	-	-	-	YES	2 (9)
C-type lectin	-	-	-	-	YES	1 (0)
Protease inhibitors						
Cystatin-type cysteine proteinase	YES	-	YES	YES	YES	3 (1)
Lipid binding						
Phosphatidylethanolamine-binding protein	-	-	YES	-	YES	1 (10)
Fatty acid retinol binding protein	YES	-	-	-	-	2 (8)
Other proteins						
TTLs	YES	-	YES	-	YES	9 (36)
Venom allergen-like proteins	YES	-	YES	-	-	0 (1)
Mucin-like protein	-	-	-	-	YES	1 (3)
Host cytokine homologues						
Macrophage migration inhibitory	-	-	YES	YES	YES	0(0)
Plant cell wall degradation enzymes						
Pectate lyase	YES	YES	-	-	-	0 (1)
Endoglucanase (and precursors)	YES	YES	-	-	-	1 (27)

References [1]- Bellafiore et al., 2008; [2]- Wang et al., 2001, Gao et al., 2003; [3]- Bennuru et al., 2003; [4]- Moreno and Geary 2008; [5] – Hewitson et al., 2008